Persuasion Dialogues & Opponent Modelling

Hadjinikolis, Christos

Awarding institution:
King's College London

Persuasion Dialogues
&
Opponent Modelling

Christos Hadjinikolis

Department of Informatics

King's College London

A thesis submitted to King's College London in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The 31st of October 2014

This thesis is lovingly dedicated to my family, whose invaluable support, encouragement, and constant love have sustained me throughout my life.

Acknowledgements

There are people in everyone's lives who make success both possible and rewarding. It would not have been possible to write this doctoral thesis without the help and support of these people, to only some of whom it is possible to give particular mention here.

First and foremost, I am indebted to Dr. Sanjay Modgil, my principal supervisor, for his patience, moral support, guidance, and inventive ideas these past four years. I feel fortunate to have him as my supervisor.

I am to also grateful to Dr. Elizabeth Black, my second supervisor, for her support as well as for her constructive feedback over the years.

I must mention my third supervisor Prof. Peter McBurney whose experience along with his kind and cheerful spirit have served many times in encouraging me to overcome the various challenges faced in a PhD, and Prof. Michael Luck, as he was the one who initially helped me take the first steps into my PhD, making sure to quickly put me on the right track.

As far as first steps are concerned, I would also like to include here Prof. Costas Iliopoulos, and my friend Dr. Solon Pissis, for their useful advice and guidance.

I want to also express my gratitude to Assistant Prof. Aspasia Daskalopoulou, my diploma thesis supervisor and friend, as she was the first person to support my decision to pursue a PhD.

Many thanks go to Dr. Antonis Bikakis, for his support and advice as well as for our stimulative conversations, and for all the opportunities he gave me to work with him.

Special thanks go to Dr. Yiannis Siantos for his unconditional friendship and support, as well as for his contribution to this thesis. His brilliant mind, and sharp wit have been of invaluable worth to me and I can only hope that our relationship will keep on growing both within as well as out of the academia.

I am also thankful to my examiners Prof. Kostas Stathis and Prof. Paul E. Dunne, for their reviews and for their useful comments.

I would like to acknowledge the financial, academic and technical support of the Department of King's College London, and its staff, particularly in the award of a partial Postgraduate Research Studentship that provided a significant financial support for this research.

Last, but not least, I would like to thank my friends Demetris, Akis and Flora for their support and encouragement, as well as my parents and sister, who have all given me their unequivocal support throughout as always, for which my mere expression of thanks likewise does not suffice.

London, October 2014

Abstract

This thesis orients around argumentative characterisations of logical non-monotonic reasoning, focusing on the arbitration between conflicting claims. These characterisations are studied in terms of *argumentation systems*. In this context, groups of inference patterns, composed of *arguments* for and against a claim, are produced and evaluated for the purpose of testing the acceptability of that claim. The objective of this thesis is to investigate the generalisation of argumentation systems to communicative (dialogical) interactions, in which the reasoning process is distributed among *opposing* agents.

Under this scope a variety of issues arise such as the form of these dialectics, the development of protocols concerned with different forms of argument evaluation, strategy development for decision making, and modelling of opponent knowledge used in strategy development.

This thesis makes two main contributions to the study of dialogues. The first is the provision of a dialogue framework for structured argumentation. Through this framework it is shown that the structural form of arguments needs to be taken into account when strategising, since it may have considerable impact on the outcome of a game. It is also shown that not accounting for the structural form of arguments may compromise the *soundness* of argument evaluation results. The second is the provision of a modelling formalism which defines how information possibly known to an opponent can be built, updated and maintained in the form of an *opponent model*. Part of the proposed modelling methodology relies on statistical inference and can find practical application both within the broad area of artificial intelligence and multi-agent systems as well as in other areas.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The word logic has its origin in the Greek word *'logike'* which refers to the ability to *reason*. The formalisation of reasoning is the study of logics. As Brewka [1991] explains, logics represent formal forms of reasoning where the study of inference is researched from a normative perspective and not through an empirical one.

This thesis orients around argumentative characterisations of reasoning in non-monotonic logics, formally expressed as *argument systems*. Such systems focus on the arbitration between conflicting conclusions in a logic, and are concerned with certain groups of inference patterns in which *arguments* for and against a claim are produced and evaluated, for the purpose of testing the tenability of that claim (Prakken and Vreeswijk [2002]). That is, to test whether that claim can be deemed *acceptable* with respect to a *semantics* (e.g. credulously or sceptically acceptable). This process of justification is referred to as *proof theory*.

A rather seminal work in this field is the abstract formalism for argumentation-based inference introduced by Dung [1995]. In Dung's work, arguments, which are typically defined with an internal structure consisting of: a set of prerequisites/premises; an inference method, and; a consequent/claim; are in contrast simply expressed as nodes in a graph which are characterised by a binary attack relation, which is in turn expressed in the form of directed edges in this graph. These two components form an argumentation framework (AF) (Figure 1.1).

In this abstract setting an argument's acceptability is evaluated under a given semantics, based on whether that argument is a member of one or more subsets of arguments in the framework, characterised by certain properties. These subsets

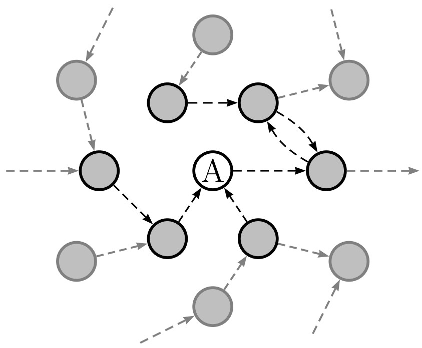


Figure 1.1: An abstract AF

are referred to as *extensions*.

Essentially, in his work, Dung was able to show that non-monotonic logics are instantiations of his theory of argumentation. According to Prakken [2010], Dung's theory was a breakthrough in how it provided a general and intuitive proof-theoretic semantics for non-monotonic logics. Arguing on the usefulness of Dung's framework Prakken explains that Dung's formalism is best seen as a tool for the analysis of argumentation systems and for the development of metatheories of them. Especially, since a consequence of the establishment of formal results for his framework is that they are inherited by its instantiations. Such results are offered in the work of Modgil and Caminada [2009] and Vreeswijk and Prakken [2000] on the development of argument game proof theories for abstract argumentation frameworks, through which one can best appreciate Dung's intuitive setting.

Argument games are procedures concerned with the evaluation of arguments in an AF which capture the human fundamental principle of *reinstatement* that characterises dialogical interactions (Caminada [2006]). An argument game is a two person game. Provided an AF, two opposing parties respectively bearing the roles of a proponent (Pr) and an opponent (Op), alternate turns exchanging arguments in a game initiated with the introduction of an argument, whose participation in a certain extension is debated. The winner of the game is decided based on "who has the last word". If Pr wins, then the participation of the disputed argument in a certain extension is justified, deeming the argument accordingly acceptable with respect to the semantics that characterise that extension.

Different protocols regulate these games, imposing restrictions on either of the participants in accordance to different semantics. The essence of the result of an argument game defined on the basis of an AF which is in turn instantiated from one's knowledge base, is found in the fact that it reflects logical inferences derived from that knowledge base. In other words, if an argument is acceptable under a given semantics in the framework, then the claim supported by that argument can also be inferred with respect to those semantics in the logic.

Due to their appealing and intuitive structure, abstract argumentation systems (AAS) serve as ideal platforms for underpinning actual dialogical interactions (dialogues). The focus of this thesis is to investigate how one can generalise argumentation to such interactions, where knowledge as well as reasoning is distributed amongst participating agents[1].

In this context a variety of issues arise. We are particularly concerned with:

- the development of a framework for persuasion dialogues;
- the locutions (logical utterances) used in them, particularly in relation to how these locutions can be used for the implicit construction of arguments;
- protocol development, for the evaluation of those arguments within the framework;
- the investigation of the strategic considerations that should characterise an agent's decision making, and;
- the construction of models of opponent information intended to be used in strategising.

1.1 Current State of the Art

In recent years, research in agent dialogues has been attracting much interest (Prakken [2006]), especially in relation to the *best response problem*. Dunne and

[1]Agents are autonomous artificial entities that rely on logics to reason and interact in uncertain environments, for the purpose of completing tasks delegated to them.

McBurney [2003] and Dunne and McBurney [2004] define this problem within the context of dialogues as one where a participant selects a locution, among all available locutions at a certain point, to introduce in a *game* which by some measure is deemed optimal. In the case of competitive contexts, "optimal" is understood in terms of increasing a participant's self-interested objectives (Amgoud and Maudet [2002]; Black and Atkinson [2011]; Black and Hunter [2009]; Carmel and Markovitch [1998]; Emele et al. [2011]; Oren and Norman [2010]; Prakken [2005]; Rienstra et al. [2013]; Riveret et al. [2007, 2008]; Rovatsos et al. [2005]; Walsh et al. [2002]). In the context of our research, moves and games are respectively translated to *arguments* and *dialogues*. Another focus is concerned with protocol design issues, usually related with the analysis of the strategical concepts that characterise a participant's choice, for the production of desirable dialogical outcomes (Black and Atkinson [2011]; Black and Hunter [2009]; Fan and Toni [2011]; Prakken [2006]; Rahwan and Larson [2008]).

Generally, research concerned with game-playing and the development of search strategies in games has evolved around von Neumann's [1928] minimax theory. The fact that in many cases researchers tend to provide variants of the minimax algorithm[1] (Carmel and Markovitch [1996]; Oren and Norman [2010]), for dealing with strategising in a variety of competitive game contexts, serves as proof of von Neumann's impact in the field. The general idea is the instantiation of a *game tree* which simulates the possible ways based on which a game may evolve. One may then evaluate this tree by assigning utility values to its possible outcomes, in order to choose the next move to make in the game.

However, application of the minimax algorithm is not unconditional as it only accounts for games of *perfect information*. In other words, if a participant is not aware of the possible legal actions/moves that its counterpart may employ at all times in a game, then the application of the minimax algorithm may be ineffective.

So as to allow for the application of von Neumann's theory, as well as for the use of other approaches for strategy development, in games of *imperfect information*, some researchers propose that a participant's strategical considerations

[1]A recursive algorithm used for choosing the next optimal move based on that move's expected utility.

be based on the anticipated outcomes of a game. Outcomes, which may result from a variety of choices and which can be evaluated in accordance to how a dialogue may evolve. To do so participants have to rely on the modelling of their interlocutors' replies to each choice. This naturally reflects real world dialogues in which we select utterances based upon what we believe are our interlocutor's beliefs (and goals). The act of modelling the interlocutors' replies at each choice point is one way of *opponent modelling*.

Basically, an opponent model (OM) consists of five basic components: an opponent's beliefs, abilities, preferences, objectives, and strategy. Numerous researchers who deal with the best response problem in dialogues rely on opponent modelling for implementing, and employing strategies, some of which are Black and Atkinson [2011]; Carmel and Markovitch [1996, 1998]; Emele et al. [2011]; Oren and Norman [2010]; Riveret et al. [2007, 2008]; Rovatsos et al. [2005]; Walsh et al. [2002]. Other researchers, e.g. Amgoud and Maudet [2002], rely on the *dialectical obligations* of a participant for implementing and employing a strategy. These are, as explained by Prakken [2006], the expectancies created by *commitments* in a dialogue, such as supporting a proposition when challenged or else retracting it. These commitments are often assumed to be accumulated in a participant's *commitment store*. Relying on such models allows one to anticipate what might follow in a dialogue, so as to optimally choose it next move.

Generally, the above approaches are successful in providing theoretically sound methods for dealing with the strategy problem or for producing desirable dialogical outcomes—outcomes characterised by some properties—by regulating the protocols of such games. However, the fact that some of them (e.g. Oren and Norman [2010]; Rahwan and Larson [2008]; Rienstra et al. [2013]; Riveret et al. [2008]) rely on abstract argumentation framework (AAS) restricts one from accounting for the logical content and structure of arguments, and their possible effects on a dialogue game, i.e. from accounting for the underlying logic.

As Prakken [2010] explains, when actual argumentation-based inference has to be modelled, Dung's framework is usually too abstract. In the context of dialogues, this limitation applies in the sense that it restricts one from accounting for how the logical contents of the arguments introduced in a debate may interact allowing for further inference. In simple words, the possibility of new information being made available as a result of the exchanged arguments in a game, eludes abstract approaches. We consider this feature to be an inherent property of the dynamic nature of dialogues. Thus, since AAS are limited with respect to this necessity, the use of structured argumentation frameworks is imperative if one is to account for this feature.

Accounting for the structure of arguments in dialogues is important in two ways. The first relates to strategising. Since argument instantiation from the logical constituents of other arguments introduced into a game is possible one may wish to adapt its strategy so as to avoid revealing crucial information to the opponent. The second relates to the development of dialogue frameworks and protocols for the production of dialogical outcomes, which can be reflected by the application of a proof theory on the accumulated knowledge, implicitly acquired from the dialogue process. If such correspondence can be shown, then these systems can be characterised as *sound* and *complete*. Not accounting for the possible instantiation of an argument in this case, may have a crucial effect on soundness and completeness claims in frameworks that rely on abstract argumentation.

An ideal framework able to serve as a dialogue development platform is proposed by Prakken [2010], who introduced an instantiation of Dung's [1995] abstract framework but for structured arguments, referred to as $ASPIC^+$. In this framework, arguments consist of defeasible/strict rules and premises, where defeat relationships resemble the notion of attacks in Dung's framework, as the term of attacks is used for defining the different kinds of contrariness between these arguments in the framework. Preferences between arguments, that derive from pre-orderings that characterise the logical constituents of arguments, are also used in the framework and can be used for deciding defeat relationships between contradictory arguments (arguments with symmetric attacks).

Other such frameworks have been used as platforms for the development of dialogue frameworks (Fan and Toni [2011]). However, only with limited exceptions (Black and Hunter [2009]; Fan and Toni [2011]; Prakken [2005]), even in approaches that rely on a structural representation of arguments, this dynamic aspect of dialogues is not explored (Riveret et al. [2007]). We should also note, that there are several proposals by which "structure" not contained within Dung's [1995] abstraction can be added such as the work of Bench-Capon et al. [2007] on *practical reasoning*[1] concerned with the use of arguments for action. The researchers state that the acceptability of an argument for an action not only bears on what is true in a given circumstances, but also on the values and aspirations of the agent to whom the argument is directed. They further argue that given that no agent can specify the relative priority of its aspirations outside of a certain context, this prioritisation must be a product of practical reasoning and cannot be used as an input to it. This requires that preferences are explicitly modelled in their framework which can also handle inconsistencies between conflicting preferences.

This is disregarded in many existing works, where only defeating arguments are moved into a dialogue, such as in Prakken [2006]; Riveret et al. [2007, 2008]. As Amgoud and Cayrol [1997] explain, the notion of defeat can be partly derived from the integration of preferences into an AF, allowing for choosing between two contradictory arguments (one attacks the other and vice-versa) if the one is preferred over the other. We note that, though some reciprocal conflicts may be resolved this way, resorting to the notion of defeat does not suggest that all such conflicts will be resolved[2]. More importantly though, in the case of dialogues, in order to rely on preferences, one has to assume that participants share the same preferences over the set of arguments used in a dialogue.

Assume for example, two $ASPIC^{+}$-arguments A and B where $A : p; p \Rightarrow q$ (where p is the premise, $p \Rightarrow q$ is a defeasible rule, and q is the claim) and $B : p; p \Rightarrow \neg q$. Since A and B support conflicting claims they contradict each other and thus share a symmetric attack relationship (particularly this is called rebut attack). While the two participants may agree that A attacks B and vice-

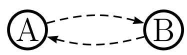


Figure 1.2: A symmetric attack relationship

versa, one may believe that B is preferred over A and so A does not defeat B,

[1] Reasoning about what to do.

[2] This is because there are weaker and strict notions of defeat (Prakken [2010]), where in the first case it may very well be that A defeats B while B defeats A at the same time.

while the other may believe that A is preferred over B and so A defeats B. However, assuming that the two participants share the same beliefs is a rather unrealistic assumption.

A useful idea would be to formalise a more expressive dialogue framework so as to allow the participants to not only argue about the acceptability of the arguments they utter, but to also support the rationale for why one argument defeats another by uttering their preferences into the game as well. The game's protocol can then regulate the circumstances under which a certain preference should be discarded over its conflicting preference or not, which could be based on the game's semantics.

A similar approach is employed in the work of Bench-Capon et al. [2007] in which the concept of "preference" is considered in terms of the notion of "audience": the relative importance that agents may ascribe to the (qualitative) social/ethical values associated with arguments. One original feature of this work being that the dialogue game described allows agents to argue over value orderings (in Amgoud and Cayrol's [1997] work preferences are rather static) so that a case for acceptance can be made (or challenged) by attacking the implied value prejudices of an opponent. The work of Amgoud and Parsons [2001] is also worth noting here, which concerns a general framework for handling dialogues between participants with different preferences not only with respect to formulæ of a logical language, but also with respect to the roles that agents may have.

In relation to the use of OMs the basic assumption is that such models can be induced from one's accumulated experience against a certain opponent (Oren and Norman [2010]; Riveret et al. [2007, 2008]). In other words, they can be produced as the sum of all the information that a modeller acquires *directly* from its opponent in their history of dialogues. Another way of inducing or extending an OM is the acquisition of information possibly provided to the modeller by a *third party*. Credibility is crucial when modelling opponent information, as strategies are developed on that information and their effectiveness is affected by it. This implies that, when modelling opponent information, different information collection methods should be associated with different confidence levels.

More importantly though, a challenging issue concerned with opponent modelling is to investigate how one can *augment* an OM. That is to answer the question: “How can I anticipate what else could an opponent know, given my current assumptions about what he knows?”. This requires that we research ways to augment/extend an OM with additional information, which do not rely on direct or third party information collection methods.

Response to this question can partly be given by how humans augment their assumptions about the possible knowledge of their interlocutors. Assume for example that a lawyer encounters an opposing colleague in a trial for the first time ever. One would expect that both lawyers are aware of the laws related to the case or that both know the reports provided to the court in relation to the crime, e.g. concerned with a murder case. Factors like the context of the debate, or access to shared information can assist in this case. More importantly though, each of the lawyer's general experience in murder cases could assist in anticipating the opposing part's strategy and possible arguments that could be employed. This could be based on whether arguments that appear in the trial process, relate to arguments which appeared in past murder cases. This makes a modeller's *general* experience an important source of information which could somehow be exploited for opponent modelling purposes.

Nevertheless, in most cases where opponent modelling is employed, the methodologies and formal procedures with which such a model may be built and updated are often either left implicit (Oren and Norman [2010]; Riveret et al. [2007, 2008]), or are only concerned with modelling some particular components of an OM (Black and Atkinson [2011]; Carmel and Markovitch [1996, 1998]; Walsh et al. [2002]). Furthermore, the basic assumption with respect to the origin of the opponent beliefs encapsulated in an OM, is that these beliefs are the collection of the distinct utterances that an opponent has put forth in dialogues with the modeller. Relying on this assumption implies that an agent's experience is exploited in a somewhat monolithic way, since an agent's accumulated dialogue experience may encode additional information which could also be used for modelling, and which could lead to an increase of the effectiveness of strategies that rely on OMs.

Let us now attempt a more thorough review of the literature with respect to the highlighted issues.

Rahwan and Larson [2008] examine the notion of strategy within argumentation focusing on showing how dialogues may be designed so as to prevent agents from hiding information or lying. Specifically, their research concerns the development of an argument mechanism that would be 'strategy proof', under specific circumstances (related to topological restrictions on an argument graph). However, they rely on an AAS and thus the impact of the underlying logic is not explored in their work.

Amgoud and Maudet [2002] explore aspects concerned with the strategy problem in the context of persuasion dialogues. Their approach relies on choosing between two kinds of strategies in a game, a 'destroy' and a 'build' strategy based on information found in a participant's commitment store. A 'build' strategy consists of moving in and defending some arguments in a player's own commitment store, while the 'destroy' strategy consists of attacking some commitments (arguments) in the opponent's commitment store. Though they do rely on a structural representation of arguments, information about the opponent is only limited to an on-the-fly acquisition of knowledge provided by opponent arguments moved in the game. Therefore, the information acquired during a game is not exploited in later dialogues. In addition, they assume only simple protocols ignoring, for example, features like *backtracking* (a feature in dialogues which allows one to use alternative responses).

Riveret et al. [2007] is concerned with the integration of a probabilistic approach into an AF in an effort to deal with issues related to uncertainty and to how probabilities can be used for utility evaluation purposes. In essence, for the purpose of their work they assume dialogues where the arguments introduced into a game are evaluated as *successful* or not by an adjudicator. This success is decided on the basis of whether an argument can be constructed (based on the chance of each of its constituents to be accepted by the adjudicator) and the existence of an accepted counter-argument. They rely on structured representations of arguments where their constituents are associated with success probabilities (the chance of these constituents to be accepted by the adjudicator). The construction chance of an argument is calculated as the propagated probability of its constituents. However, construction chance of an argument composed of information provided by both participants in the dialogue is not explored.

Riveret et al. [2008] deal with determining *optimal strategies* in argument-based dialogue games, similarly relating probabilities with arguments. In their work, a strategy qualifies as optimal with respect to the probability of the participant's arguments to be accepted and the cost/benefit of those arguments. They also assume the context of adjudication debates, though in contrast to Riveret et al. [2007], their work relies on an AAS and thus the possible impact of the structural elements of arguments is not explored in their work.

As far as opponent modelling is concerned, both Riveret et al. [2007] and Riveret et al. [2008] model the possible knowledge of their opponents in the form of arguments, assuming that arguers are perfectly informed about all the arguments previously advanced by all other players, i.e., based on their interaction history. Their investigation as well concerns *games of perfect information*, and assumes that the participants' goals always comply with the dialogical objectives of the game; an assumption which, as McBurney and Parsons [2002] argue, is not always valid. Furthermore, no formal methodology is provided with respect to how such a model may be obtained. In addition, both approaches as well as the approach by Prakken [2006] rely on the notion of defeat, which requires that the participants in dialogues share the same preferences over the arguments they utter and which, as it has been stressed is a rather unrealistic assumption.

Oren and Norman [2010] are concerned with strategy development with respect to the best response problem based on OMs. They present a generally complete approach through modelling both an agent's knowledge in the form of arguments as well as their goals, while they allow for nested opponent models. The latter is a recursive modelling structure[1] that allows one to model the possibility that the opponent could also have a model of the modeller in his knowledge base, allowing for multiple levels of nesting (Carmel and Markovitch [1996]). However, once again, no information is provided with respect to how this model may be built and updated. In addition, they also rely on an AAS and thus they also do not account for the impact of the logic in strategising.

Rienstra et al. [2013] also rely on OMs, and extend the work of Oren and Norman [2010] through proposing a categorisation of three kinds of OMs. A simple OM, an uncertain OM and an extended OM differentiating between levels of

[1] i.e.: A zero level nested model is one that only accommodates one's opponent evaluation function, while a first level nested model additionally accommodates the opponent's opponent model of the first.

uncertainty of the modelled information between the tree models. They propose a rather interesting notion of modelled knowledge to which they refer as *virtual arguments*. These are arguments that a modeller thinks the opponent could be aware of but does not exactly know what they support or their exact structure. They also show how OMs can be updated in their framework while they test their work, with respect to anticipating opponent knowledge, using a random abstract argumentation theory (as big as 10 arguments). Their results focus on abstract argument, though they state that they intend to adjust their heuristics so as to account for structured argumentation.

In some cases, the fact that the formalisation of OMs is left implicit can be attributed to the boundaries of one's research scope. For example it could be that the proposed heuristics of a particular work are concerned with explicit research contexts where modelling additional information would be redundant or where various aspects of the concerned problems are purposely simplified, so as to make analysing them feasible. For instance, Carmel and Markovitch [1998] are simply concerned with modelling an opponent's strategy, expressed as a deterministic finite automaton, which essentially models a finite set of actions concerned with the iterative version of the prisoner's dilemma game. They are thus indifferent to a participant's possible beliefs. Similarly, Walsh et al. [2002] assume a finite set of predefined strategies that an agent may employ for analysing complex strategic interactions with respect to stability and game theoretic equilibria.

In another work Carmel and Markovitch [1996], again, are concerned with providing a variant of the minimax algorithm based again on modelling an opponent's strategy, this time as an evaluation function, though the provision of a recursive model that is able to define complex levels of nested opponent models. However, they provide no means based on which such a model may be acquired, while additionally one could argue that the generalised approach they propose can be applied only in contexts where the set of the possible actions that may follow from a certain game state is finite and known to all participants.

Some interesting exceptions are offered in the literature with respect to modelling. One such is proposed by Black and Atkinson [2011] and concerns a mechanism that enables agents to particularly model *preference* information about others—what is important to another agent—and then rely on it for making proposals that are more likely to be agreeable. In this case the mechanism responsible for modelling an agent's preferences is explicitly provided. Also the work of Rovatsos et al. [2005] is worth mentioning, which explores how an agent can learn stereotypical sequences of utterances in dialogues for deducing an opponent's strategy, though they do not rely on OMs. We finally note the work of Emele et al. [2011] who explore the development of an OM based on what norms or expectations an opponent might have, which, however, is not concerned with an opponent's general beliefs.

1.2 Research Questions & Contributions

Through this introduction we have identified a number of issues related to the study of argument-based dialogues. With respect to these issues the aim of this thesis is to research the following questions:

- **In relation to accounting for the underlying logic:**

  1. What is the impact of the underlying logic in an agent's strategising, and to what extent should it be accounted for strategy development purposes?
  2. If its impact is crucial, then can we define a dialogue framework on the basis of an argumentative system able to account for the structural form of arguments?
  3. How will the employment of a structured argumentation system add to the expressiveness of dialogues produced in this framework, allowing for the introduction of a participant's preferences as a means of justifying their rational with respect to a defeat relationship?

  As a result of researching these questions we further investigate:

  4. How can we evaluate the dialogue outcomes produced by our framework with respect to the acceptability of a disputed argument?
  5. With what restrictions should protocols developed for our dialogue framework be characterised with, so as to guarantee the soundness of a dialogue's result with respect to certain semantics?

- **In relation to the issue of opponent modelling:**

  6. What are the main mechanisms based on which one may build, update and maintain an OM?
  7. What factors affect the credibility of acquired opponent information?
  8. Can we formalise an opponent modelling process based on which one could account for these factors in a way that could possibly increase the effectiveness of a strategy?
  9. Can we develop a modelling mechanism for inducing and augmenting an OM based on a modeller's general experience in dialogues?

  As a result of researching these questions we further investigate:

  10. If the proposed modelling process is computationally demanding then can approximative approaches be used instead?
  11. How can we evaluate the effectiveness of an opponent modelling mechanism in increasing a model's validity and consequently its credibility?

Through responding to these questions we make two main contributions to the study of dialogue games. The first is the provision of an $ASPIC^+$-based dialogue framework, which focusses on persuasion dialogues. We chose $ASPIC^+$ for three reasons. Firstly, because it is an expressive framework which explicitly models the logical content and structure of arguments, an essential requirement for the objective of our research. Secondly, because it accommodates many existing logical approaches to argumentation (Modgil and Prakken [2011]). Finally, because $ASPIC^+$ has been shown to satisfy Caminada's [2007] rationality postulates[1].

We add to the framework's expressiveness by allowing participants to utter preferences and not just arguments in dialogues, enabling participants to 'argue' about the defeat relations of arguments they use. Relying on it, we research logical, modelling and strategic aspects, while we particularly investigate the importance of accounting for the underling logic when strategising. We show that strategies that don't account for the structural form of arguments are strategically sound only at an abstract level. In addition, we provide 2 protocols with respect to

[1]Closeness and consistency conditions.

a credulous and a sceptical semantics and prove the soundness and completeness of dialogues produced in our framework.

The second is the provision of a modelling framework for building, updating and maintaining an OM. The framework allows the modeller to differentiate between 3 ways of opponent information acquisition. Namely these are, direct collection of information from dialogues, third party provided information, and information acquired through the method of augmentation. The model assumes a structured representation of knowledge in the form of logical constituents for argument instantiation. Depending on the information acquisition method, these constituents are associated with a *confidence value* which can be used for the purpose of increasing the effectiveness of the modeller's strategy. Focussing on the augmentation process we also develop a mechanism which allows a modeller to incorporate additional information (external to the model) from its general experience into an OM, after relating it with information already in the model, i.e. to *augment* it, utilising its experience in a multifaceted way.

To give the reader an indication of the utility of the work in this thesis, we note that in its most general sense this work deals with logics for reasoning in the presence of uncertainty and conflict. Our interest is to enable agents to reason using these logics, but in a distributed manner. Examples of applications where distributed reasoning is important are described by Modgil et al. [2013] and concern the fields of philosophy, communication studies, linguistics, psychology and artificial intelligence. The work of Morge et al. [2013] is also worth mentioning, which is concerned with an argumentation-based agent model able to support service and partner selection in service-oriented computing settings. We should also note the work of Fan et al. [2013] who propose a decision making model which is applied to clinical trial selection, where given properties of a patient relevant papers are selected from a given set, that best match the patient's preferences/properties. A specific example is presented in the work of Tolchinsky et al. [2012], on deliberation dialogues about safety, where agents need to debate whether a diagnosis is correct. We note that in any kind of dialogue there may be points at which agents have to resolve disagreements and thus need to turn to persuasion.

As to the broader contribution of our work, especially in relation to the proposed modelling framework, we note that in essence what we present is a prediction mechanism. One that relies on an agent's experience for anticipating its opponents possible moves in a dialogue game. Apart from the application of the proposed approach in the context of dialogues, this mechanism could easily be generalised through appropriate modifications and applied in other A.I. contexts. S this will become clearer in Chapter 4, all that is necessary is that one identifies links between potentially associated information which can be quantified in accordance to how various factors of a concerned system affect these links.

Such contexts may concern the field of recommender systems, where this mechanism could find potential use in predicting items of interest to a system user such as contacts, music, books, purchase items, or even news based on their preferences and on the preferences of their peers. Other fields could be *bio-informatics*, for predicting whether certain amino-acids could interact to form a new protein. In this sense the impact of our contribution extends beyond the A.I. field as practical applications can be developed and applied for the solution of problems outside the scope of the A.I. field.

1.3 Thesis Structure

The thesis is structured as follows:

- Chapter 2—provides the necessary theoretical background on which we build our work.
- Chapter 3—presents the general dialogue framework developed in this thesis, and discusses the impact of the underlying logic in strategising. It also provides a formal proof of the soundness and completeness of our approach in producing dialogue outcomes that reflect certain acceptability semantics.
- Chapter 4—discusses the notion of opponent modelling, and proposes a modelling framework for building and updating an OM, as well as for augmenting it with additional content based on the relationship of information contained in it with information external to the model.

We show how confidence values may be associated with the structural logical information contained in an OM to be used for strategising purposes. At the same time, we focus on the *augmentation mechanism* showing how it accounts for additional knowledge *likely* to be known to one's opponent. We discuss the tractability of the mechanism, in relation to which we propose an approximative approach for the computation of this *likelihood* which relies on *Monte-Carlo simulations*, and test our approach.

- Chapter 5—extends the opponent modelling framework presented in the previous chapter, focussing on accounting for a number of shortcomings of the proposed augmentation process through a number of extensions, while possible modelling variations also concerned with augmentation purposes are discussed and proposed.

- Chapter 6—provides a general mechanism for strategising which utilises confidence values through a utility evaluation function (UEF) and the application of the minimax algorithm on a game tree which sums all the possible terminal states of a simulated dialogue game, illustrating how these confidence values can be utilised.

- Chapter 7—presents a methodology towards evaluating our augmentation approach in relation to whether it can positively affect the credibility of an OM, given the absence of argument-based benchmarks for agent discourse.

- Lastly, Chapter 8—offers a short summary of our research, listing our contributions, while it poses our future research objectives on issues which arise from our work.

1.4 Publications

We note that some contents of this thesis also appear in the following publications:

- Hadjinikolis et al. [2012a]

  Christos Hadjinikolis, Sanjay Modgil, Elizabeth Black, Peter McBurney, Michael Luck. *Investigating Strategic Considerations in Persuasion Dia-*

logue Games, Proceedings of the Sixth Starting AI Researchers' Symposium (STAIRS), 2012, IOS Press.

- Hadjinikolis et al. [2012b]

  Christos Hadjinikolis, Sanjay Modgil, Elizabeth Black, Peter McBurney. *Mechanisms for Opponent Modelling*, Imperial College Computing Student Workshop (ICCSW), 2012, OpenAccess Series in Informatics, Schloss Dagstuhl.

- Hadjinikolis et al. [2013]

  Christos Hadjinikolis, Yiannis Siantos, Sanjay Modgil, Elizabeth Black, Peter McBurney. *Opponent Modelling in Persuasion Dialogues*, In: F. Rossi (Editor): Proceeedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI), 2013, Beijing, China.

The first publication relates to the content presented in Chapter 3. A framework for argumentation-based persuasion dialogues is presented which allows the implementation of strategies based on a participant's modelling of its interlocutors knowledge. The framework is defined on the basis of $ASPIC^+$. Existing works on persuasion is extended (Prakken [2006]), by accounting for both admissible and grounded semantics, and also by allowing participants to not only move arguments that attack those of their interlocutor, but also preferences which undermine the success of these attacks as defeats. Formal results for these dialogues are stated, while it is also illustrated through an example that appropriate mechanisms for strategising need to account for the logical content of arguments, rather than just rely on their abstract specification.

The following two relate to Chapter 4. The first is a preliminary paper workshop paper, which introduces the ideas that follow in the second one. Main assumptions and approaches to opponent modelling are introduced and discussed. We focus on the fact that an agents experience may encode additional information which if appropriately used could increase a strategys efficiency. Rely on a modeller's experience formally expressed as its history of dialogue interactions, we define a mechanism for augmenting an OM with information likely to be dialectally related to information already contained in it. Precise computation of this likelihood is exponential in the volume of related information. Thus, we describe and evaluate an approximate approach for computing these likelihoods based on a Monte-Carlo simulation.

Chapter 2

Theoretical Background

In this chapter we provide the necessary theoretical background for our research. The chapter is structured as follows: in Section 2.1 we present Dung's [1995] work on argumentation theory; in Section 2.1.1 issues related to the problem of argument evaluation are presented, focussing on an evaluation approach referred to as argument games; in Section 2.2 we discuss the added value of argumentation theory; in Section 2.3 we present Prakken's [2010] instantiation of Dung's [1995] framework for structured argumentation on which we base the development of the dialogue framework presented in Chapter 3; in Section 2.4 we present the types of dialogues between agents researched in the literature, as well as the necessary components of a dialogue framework, and discuss the generalisation of argument games. A brief summary of this chapter is offered in Section 2.5.

2.1 Argumentation Theory

Dung [1995] builds his theory of argumentation based on the notion of an argumentation framework (AF). Essentially, a Dung argumentation framework (DAF) is a directed graph $(\mathcal{A}, \texttt{attacks})$ where $\mathcal{A}$ is the set of arguments and $\texttt{attacks}$ is a binary relation on $\mathcal{A}$. Assume for example two arguments $A, B \in \mathcal{A}$ then if $(A, B) \in \texttt{attacks}$ then we say that A attacks B.

Definition 1 (Argumentation Framework (Dung [1995])) *An argumentation framework is a pair:*

$$AF = \langle \mathcal{A}, \texttt{attacks} \rangle$$

where $\mathcal{A}$ is the set of arguments, and $\texttt{attacks}$ is a binary relation on $\mathcal{A}$, i.e. $\texttt{attacks} \subseteq \mathcal{A} \times \mathcal{A}$.

Given a logic $\mathcal{L}$, in order to instantiate a AF from it, one needs to express both the notion of an argument as well as the notion of attack with respect to this logic. That is to define what is an argument in $\mathcal{L}$ and what constitutes an attack between two arguments. In this respect, given a set of well formed formulæ (wff) in a knowledge base Δ in $\mathcal{L}$, one can instantiate all the arguments from Δ and relate them with attack relationships.

2.1.1 Argument Evaluation

Given an AF one can attempt an evaluation of an argument's acceptability status. That is to justify whether one should or should not accept an argument. For this purpose, Dung [1995] proposes a number of acceptability semantics for his framework, concerned with different levels or forms of acceptability, whose purpose is to specify the conditions under which an argument is acceptable. These conditions, focus on a number of properties that characterise certain subsets of arguments in an AF referred to as *extensions* of the framework. The acceptability status of an argument can then be evaluated with respect to whether that argument participates in one or more of these extensions.

One of the most essential of properties that characterises all of the extensions of an AF is that of conflict-freeness.

Definition 2 (Conflict-freeness (Dung [1995])) *A set of arguments $S \subseteq \mathcal{A}$ is said to be conflict-free if there are no arguments A and B in S, such that:*

$$(A, B) \in \texttt{attacks}$$

Based on this property Dung defines a basic extension referred to as the *admissible* extension.

Definition 3 (Admissible Extension (Dung [1995]))

1. *An argument $A \in \mathcal{A}$ is acceptable with respect to a set S of arguments iff for each argument $B \in \mathcal{A}$: if $\exists(B, A) \in$ `attacks` then B is attacked by S*[1].

2. *A conflict-free set of arguments S is admissible iff each argument in S is acceptable with respect to S.*

A number of extensions are then defined in Dung's paper, which are essentially special forms of the admissible extension. These are the *complete*, *preferred*, *stable* and the *grounded* extensions. The preferred and the grounded extensions are special forms of the complete extension. Other forms of semantics, with corresponding extensions, are offered in the literature such as the semi-stable semantics or the stage semantics (Baroni and Giacomin [2009]). As in this thesis we are concerned with notions of credulous and sceptical acceptability in the way that these are reflected by the preferred and the grounded semantics respectively[2], we only present these two semantics, as well as the complete semantics which relate to them, in the way they are defined in Dung's [1995] paper. Respectively, these semantics represent the cases where an agent either chooses to credulously accept the validity of an argument, or accept it beyond any doubt (grounded).

Definition 4 (Complete Extension (Dung [1995])) *An admissible set S of arguments is called a complete extension iff each argument, which is acceptable with respect to S, belongs to S.*

Definition 5 (Preferred Extension (Dung [1995])) *A preferred extension of an AF is a maximal (with respect to set inclusion) complete extension of an AF*[3].

Definition 6 (Grounded Extension (Dung [1995])) *The grounded extension of an AF is* the *smallest complete extension in AF.*

[1]Specifically this means that $\exists C \in S$ such that $\exists(C, B) \in$ `attacks`.

[2]Apart from the the grounded extension, sceptical justification of an argument can be based on its membership in all the preferred extensions of an AF (Vreeswijk and Prakken [2000]).

[3]There may be more than one maximal complete extensions in an AF.

2.1.2 Argument Game Proof Theories

Though in his work Dung [1995] defined the properties that argument extensions should have in order to correspond to a semantics, the heuristics for their identification and construction were not provided[1]. Modgil and Caminada's [2009] argument game proof theories address this problem.

Generally, argument games are particularly concerned with the justification of a single argument in a AF with respect to a certain semantics. These games build on the reinstatement principle since an argument is deemed acceptable if it can adequately defend itself against attacking arguments or be supported against them by other arguments in the framework.

Similar approaches related to the development of proof theories for argumentation can be found in the literature concerned with a number of semantics, such as Caminada's [2006; 2007] *labelling*, and the argument games proposed by Vreeswijk and Prakken [2000]. However, we focus on the work of Modgil and Caminada [2009], as it mostly relates to the research of this thesis.

Intuitively, as Modgil and Caminada [2009] explain, the idea of argument games assumes the moving of arguments as in two-person dialogue games, providing a natural way in which to lay out and understand the heuristics for these proof theories. The game consists of two participants, a proponent (Pr) and an opponent (Op) who move arguments against each other in turns. This process is called *dispute*. A dispute is won by the one who has the last move and thus reflects the reinstatement principle.

Take for example the AF in Figure 2.1a and assume an argument game for the acceptability of argument A. A_1 would be moved first by Pr to be countered by Op with B_2. At this point Pr is presented with a choice of moves: C and D (Figure 2.1b). This implies that from this point and on two distinct disputes can evolve. This can be expressed in the form of branching in a tree referred to as *dispute tree*, which is defined as follows:

Definition 7 (Dispute Tree (Modgil and Caminada [2009])) *Let AF be an argumentation framework* $AF = \langle \mathcal{A}, \texttt{attacks} \rangle$ *, and let* $A \in \mathcal{A}$*. The dispute tree*

[1]Excluding the constructive method for forming the grounded extension as the fixed-point of the characteristic function F, beginning from $F(\varnothing)$ Dung [1995]

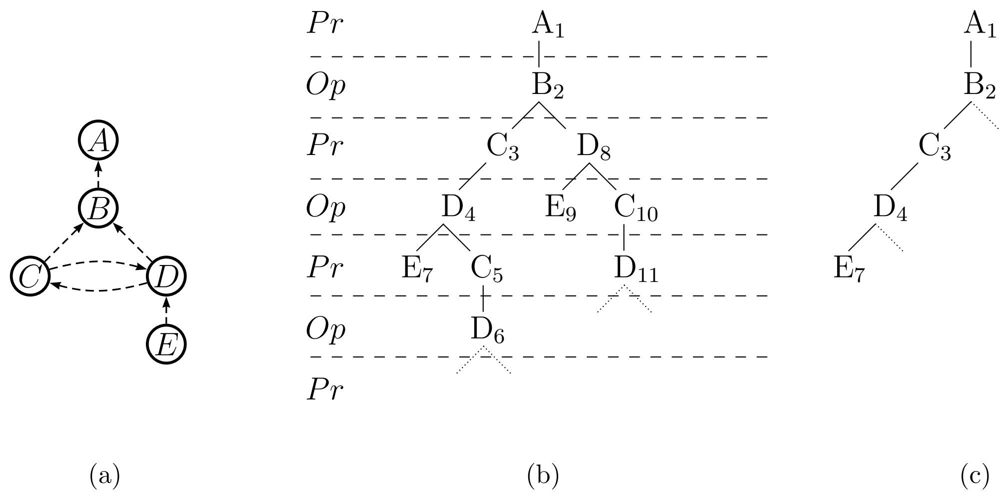


Figure 2.1: Grounded semantics: a) The AF, b) Part of the dispute tree $\mathcal{T}$, c) The winning strategy $\mathcal{T}' \in \mathcal{T}$ (Figures taken from Modgil and Caminada [2009])

induced by A in $\mathcal{A}$ is a tree $\mathcal{T}$ of arguments, such that $\mathcal{T}$'s root node is A, and $\forall X, Y \in \mathcal{A} : X$ is a child of Y in $\mathcal{T}$ iff $(X, Y) \in$ `attacks`.

Let us continue the interaction by assuming that Pr opts for C_3 which can only be countered by Op with D_4. At this point Pr is again presented with a choice, that is to either repeat C, or counter D with the use of E. Here we encounter one of the main problems that needs to be regulated in argument games, since if *repetition* is available to both participants at all times then this would result in an infinite dispute.

Let us further extend the dispute and assume that Pr opts for C_5 again, to which Op responds with repeating the same move, which is D_6. At this point though Pr is presented again with a choice between 2 moves—C or E—she is more flexible with *where* to introduce them in the dispute tree. For example, she can directly counter D_6 with either of them, or counter D_4 with introducing E_7. The latter choice is called *backtracking*. Notice that given backtracking, D is also available to Pr as a backtrack move against B_2—that is D_8 in Figure 2.1b. Also notice that even though D is an argument that has already appeared in the game by Op (D_4) it is now used by Pr.

In summary, the aspects regulated by the protocol rules of a game are:

- the repetition of arguments;
- when to allow backtracking; and
- whether Pr should reuse arguments introduced by Op and vice-versa;

Different sets of rules for these games can capture the different semantics under which an argument can be justified. In the following sections we present the rules developed by Modgil and Caminada [2009], particularly in relation to the credulous preferred semantics and the sceptical grounded semantics.

Finally, notice that, in contrast to a case where a game would consist of a single dispute, in the case of a tree structure deciding on the winner of the game is a more complex issue, since it is not necessary that a participant wins all disputes in order to win a game. For this reason, Modgil and Caminada [2009] turn to the notion of *winning strategy*:

Definition 8 (Winning Strategy) *Given an AF and a tree induced by an argument A then a winning strategy for A is a non-empty finite sub-tree T' which satisfies the following criteria:*

1. *All its paths must be finite and they must end with a Pr's move, and;*
2. *For every sub-path d' of its paths that ends with a Pr's move X, if there exists an argument Y such that Y attacks X, then there exists another sub-path d'' such that the path created by d' and Y is contained in d''.*

Specifically, in the case of the argument game of our example, a winning strategy is identified in the dispute $A_1 - B_2 - C_3 - D_4 - E_7$ (Figure 2.1c).

2.1.2.1 An Argument Game for Credulous Semantics

For their argument games Modgil and Caminada [2009] define a legal move function ϕ for distinguishing between legal and illegal moves in order to encode the rule restrictions of different argument games for a semantics S. In general, assuming that one wishes to show that X is a member of an extension E under the semantics S, the associated legal move function ϕ for S, prunes the dispute-tree induced for that argument to provide a sub-tree $\mathcal{T}'$ called ϕ-*dispute tree* that is

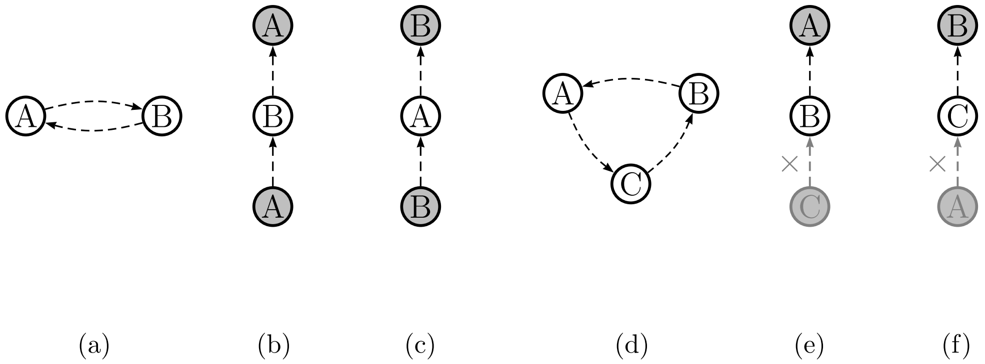


Figure 2.2: a) An AF with an even-loop between two arguments, b) An argument game for A, c) An argument game for B, d) An AF with an odd-loop between three arguments, e) An argument game for A, and f) An argument game for B

the play-field of the argument game. X can be deemed accepted in this setting, iff, there exists a *ø-winning strategy* $\mathcal{T}''$ which is a sub-tree of $\mathcal{T}'$ such that the arguments moved by Pr do not attack each other while Pr wins all disputes in $\mathcal{T}''$. For convenience, we refer to the play-field sub-tree as $\mathcal{T}$ and to a winning strategy in $\mathcal{T}$ as $\mathcal{T}'$.

For dealing with issues related to the appearance of infinite disputes, resulting from possible loops in an AF, Modgil and Caminada [2009] simply prevent one of the two participants from repeating a move in a dispute. For the credulous *preferred* semantics that participant is Op. Intuitively, this is because credulous acceptability is a more lenient form of acceptability where Pr, in effect, need only present an admissible set of arguments whereas Op has to eliminate all possibilities[1].

In a trivial case, assume an AF composed of only two arguments A and B which both attack each other (Figure 2.2a). In this case we identify two preferred extensions one containing A and one containing B. The respective argument games where Pr would try to justify either of them should end with the repetition of A and B deeming Pr the winner in each case (Figures 2.2b & 2.2b). Of course, in a case of an odd-loop, e.g. Figure 2.2d, Pr must be restricted from using arguments that attack (or are attacked by) arguments she previously introduced

[1]That this elimination process can genuinely result in infeasible computational demands is one of the consequences of the analysis of Dunne and Bench-Capon [2003].

into the game, i.e. using C or A respectively in Figures 2.2e & 2.2f.

In relation to backtracking, Modgil and Caminada [2009] explain, that it should be allowed to both participants in a justification procedure, since either bearing the burden of fully proving the acceptance of an argument or bearing that of proving it wrong, both sides should be allowed to exhaustively use all possible alternatives in order to counter their opponent. In this respect, backtracking is implicitly captured through the construction of all possible disputes that can be deduced for an argument A, through a dispute tree. Hence the legal move functions ϕ they provide are concerned with restrictions imposed on a single dispute.

Specifically, in relation to the credulous preferred games, the legal move function ϕ_{PC_1} they defined is as follows:

Definition 9 (Legal move function ϕ_{PC_1} (Modgil and Caminada [2009])) *Given an $AF = \langle \mathcal{A}, \Re \rangle$ and a dispute* d *such that X is the last argument in* d*, and $Op(d)$ the arguments moved by Op in* d*, then ϕ_{PC_1} is a legal move function such that:*

- *if d is of odd length (next move is by Op) then: $\phi_{PC_1}(d)$=\{Y |*
  - $Y \Re X$
  - $Y \notin Op(d)$
- *if d is of even length (next move is by Pr) then: $\phi_{PC_1}(d)$=\{Y |*
  - $Y \Re X$
  - $Y \in POSS(d)\}$

where $POSS(d) = \{Y \mid \neg(Y \Re Y)$ and $\forall Z \in Pr(d), \neg(Z \Re Y)$ and $\neg(Y \Re Z)\}$

Notice that Op is restricted from repeating a move she previously used while Pr is not. The use of $Y \in POSS(d)$ imposes a couple of restrictions on Pr for the purpose of devising a more efficient game. Since the arguments moved by Pr in a winning strategy are required to be conict free (essentially Pr attempts to build an admissible extension which is by definition conflict-free), shorter proofs can be obtained by preventing Pr from moving arguments in a dispute that either attack themselves or attack or are attacked by arguments that Pr has already introduced in d.

Through this set of rules—the legal move function ϕ_{PC_1})—Modgil and Caminada [2009] show that given a finite AF, there exists a ϕ_{PC_1}-*winning strategy* $\mathcal{T}'$ for X such that the set $Pr(\mathcal{T}')$ of arguments moved by Pr in $\mathcal{T}'$ is conflict-free, iff X is in an admissible extension of the AF.

2.1.2.2 An Argument Game for Grounded Semantics

The ϕ-*legal move* function for the grounded semantics proof game defined by Modgil and Caminada [2009] is very similar to ϕ_{PC_1} with the main difference that in this case, instead of Op, Pr is the one restricted from repeating her moves in a dispute. Intuitively, this is because sceptical acceptance adds a heavier burden to the proponent (in the trivial example of FIgure 2.2a, Pr should be prevented from repeating either A or B since the grounded extension for that AF is empty).

The legal move function for the grounded semantics is as follows:

Definition 10 (Legal move function ϕ_{G_1} (Modgil and Caminada [2009])) *Given an AF, a dispute d such that X is the last argument of d, and $Pr(d)$ the arguments moved by Pr in d, then ϕ_{G_1} is a legal move function such that:*

- *if d is of odd length (next move is by Op) then:*

  $\phi_{G_1}(d) = \{Y \mid Y \Re X\}$

- *if d is of even length (next move is by Pr) then:*

  $\phi_{G_1}(d) = \{Y \mid$

  - $Y \Re X$
  - $Y \notin Pr(d)\}$

In similar sense to the credulous preferred semantics, Modgil and Caminada [2009] show that assuming that AF is finite, then there exists a ϕ_{G_1}-winning strategy $\mathcal{T}'$ for X such that the set $Pr(\mathcal{T}')$ of arguments moved by Pr in $\mathcal{T}'$ is conflict-free, iff X is in the grounded extension of AF.

2.2 The Added Value of Argumentation Theory

The essential contribution of Dung's [1995] work was that he showed that his argumentation theory is an abstraction of non-monotonic logics. Non-monotonic logics, such as logic programming or default logic theory, are formal logics where the addition of a formula to a theory may produce a reduction of its set of consequences (Brewka [1991]). In contrast to monotonic logics, non-monotonic logics are closer to our common-sense reasoning since we rely on assumptions for inferring a conclusion which we may retract in the face of new information. Such an assumption is *negation as failure* employed in `Prolog`—a logic programming language—where a negative conclusion is inferred upon the absence of information based on which an affirmative conclusion would be derived, e.g.:

`NOT P` is true whenever `P` cannot be derived

In these logics the justification of a formula is based on whether that formula can be entailed from a proof theory, e.g.:

$$\Delta_{NML} \vdash \alpha$$

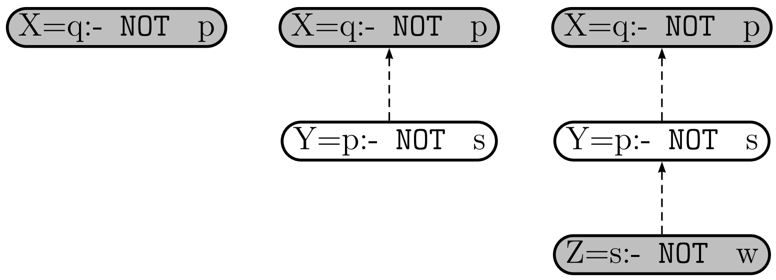
That is, given a Δ in a non-monotonic logic entail α by applying a proof theory. Specifically in relation to default logics, these theories are called *default theories* and they are related with the notion of extensions. Each extension is interpreted as an acceptable set of beliefs, and choosing among these possible extensions assumes the employment of techniques associated with different acceptability semantics (Reiter [1980]).

Dung was able to show that given a Δ one can instantiate an AF and apply a proof theory on the framework, instead of Δ, in order to entail/justify α, i.e.:

$$AF \vdash \alpha$$

At the same time he defined a number of extensions of arguments in his framework which correspond to extensions of beliefs in a non-monotonic logic.

As Prakken [2010] explains, Dung's [1995] article was a breakthrough in the



(a) (b) (c)

Figure 2.3: Distributed reasoning between two agents Ag_1 (grey) and Ag_2 in logic programming.

way that it provided a general semantics for the consequence notions of argumentation logics, while it allowed for different systems to be categorically compared after being translated into his high level format. In consequence, his work has been rather influential and resulted in giving an enormous boost to research in computational argumentation. However, Modgil et al. [2013] argue that its success is mostly attributed to the fact that his framework is able to serve as a basis for the definition of proof theories for distributed non-monotonic reasoning, simply based on the *reinstatement principle*. A principle which very much appeals to our human intuition.

For example, imagine an argument X in logic programming where:

$$X = [q :- \ \texttt{NOT} \ \ p]$$

is 'uttered' by an agent Ag_1 within the context of a dialectical interaction with another agent Ag_2 (Figure 2.3a). So far we can assume that the consequent q is justified. Further assume that Ag_2 utters an argument:

$$Y : [p :- \ \texttt{NOT} \ \ s]$$

(Figure 2.3b). Now, given that p is produced as the consequent of Y, then NOT p can no longer hold and thus q is no longer justified. However, it is possible for Ag_1 to *reinstate* the acceptability of X, consequently justifying q, by uttering an argument:

$$Z : [s :- \texttt{NOT}\ w]$$

(Figure 2.3c). Since the employment of an AF allows the modularisation of knowledge in the form of arguments then this example suggests that rather than applying a proof theory directly in logic programming, one can reason about the justified formulæ through the implicit and incremental construction of a Δ through the exchange of arguments between two opposing agents.

As Modgil and Caminada [2009] state, the argument game approach, places an emphasis on the dialectical nature of argumentation as it appeals more to an inter-subjective notion of truth:

> *"Truth becomes that which can be defended in a rational exchange and evaluation of interacting arguments."*

In this sense, dialectical semantics allow one to easily relate formal entailment to discussions and debates. Since this approach is akin to human models of reasoning, it allows for the integration of human reasoning in dialogues, enabling human to human interactions, human to computational interactions and computational to computational interactions. Therefore, apart from these argument games serving as guidelines and principles for the design of algorithms, they also assist in bridging the gap between human and computational reasoning.

2.3 The $ASPIC^+$ Framework

Prakken [2010] argues that Dung's [1995] work has been very successful, especially if seen as a tool for analysing particular argumentation systems and for the development of a meta-theory for them. However, it has been criticised for being too abstract when actual argumentation-based inference has to be modelled (Prakken [2010]), which highlights the need for a structural instantiation of his approach.

In his work Prakken [2010] instantiates Dung's abstract setting, accounting for the structure of arguments and their *defeat* relation. Arguments are presented in the form of inference trees, relying on strict and defeasible rules, while three ways of attacks[1] are defined.

In essence, Prakken builds his *ASPIC*$^+$ framework by extending the European ASPIC project (Amgoud et al. [2006]), which added expressiveness to Dung's abstract formalism, and which gave rise to further work by Caminada and Amgoud [2007] who defined a number of rationality postulates criticising some specific rule-based argumentations systems for not satisfying them. Prakken explains that for this purpose only a simplified version of ASPIC was needed. As a result, notions like *preferences* and the concept of a *knowledge-base* were not considered.

In his work the ASPIC framework is extended in 4 basic ways:

1. A third notion of attack between arguments is introduced, inspired by the work of Vreeswijk [1993] (the other two are undercut and rebut attacks);

2. Drawing from the work of Bondarenko et al. [1997], and Verheij [2003], a contrariness relation is defined to characterise the three notions of attack;

3. Premises are distinguished in 4 kinds, inspired by Gordon et al. [2007];

4. Attack relations are resolved through defining preference orderings on arguments, derived from orderings on their structural elements, and shows how *ASPIC*$^+$ satisfies Caminada's [2007] postulates[2].

We briefly present *ASPIC*$^+$ in the next section, on which we rely for the development of the general dialogue framework presented in Chapter 3.

2.3.1 Basic Definitions

Prakken's 2010 *ASPIC*$^+$ instantiates Dung's abstract approach by assuming an unspecified logical language $\mathcal{L}$, and by defining arguments as inference trees formed by applying strict or defeasible inference rules of the form $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ and $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$, interpreted as:

[1]In Prakken's framework, 'defeats' rather than 'attacks' resemble the notion of attacks in a Dung-*AF*. Attacks in *ASPIC*$^+$ are reserved for defining the contrary/contradictory relationships between the arguments in an *AF*.

[2]Though *ASPIC* also accommodates preferences satisfaction of the rationality postulates was not shown for them.

"if the antecedents $\varphi_1, \dots, \varphi_n$ *hold, then* without exception*, respectively* presumably*, the* consequent φ *holds."*

To define attacks, minimal assumptions on $\mathcal{L}$ are made; namely that certain well formed fomulæ(wff) are a contrary or contradictory of certain other wff. Apart from this the framework is still abstract: it applies to any set of strict and defeasible inference rules, and to any logical language with a defined contrary relation.

The basic notion of $ASPIC^+$ is an argumentation system.

Definition 11 (Argumentation system) *Let* $AS = (\mathcal{L}, {}^-, \mathcal{R}, \leq)$ *be an argumentation system where:*

- $\mathcal{L}$ *is a logical language.*
- ${}^-$ *is a contrariness function from* $\mathcal{L}$ *to* $2^{\mathcal{L}}$*, such that:*
  - φ *is a* contrary *of* ψ *if* $\varphi \in \overline{\psi}$*,* $\psi \notin \overline{\varphi}$
  - φ *is a* contradictory *of* ψ *(denoted by '*$\varphi = -\psi$*'), if* $\varphi \in \overline{\psi}$*,* $\psi \in \overline{\varphi}$
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ *is a set of strict* $(\mathcal{R}_s)$ *and defeasible* $(\mathcal{R}_d)$ *inference rules such that* $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$*.*
- $\leq$ *is a pre-ordering on* $\mathcal{R}_d$*.*

Arguments are then constructed with respect to a knowledge base that is assumed to contain three kinds of formulæ.

Definition 12 (Knowledge-base) *A knowledge base in an argumentation system* $(\mathcal{L}, {}^-, \mathcal{R}, \leq)$ *is a pair* $(\mathcal{K}, \leq')$ *where* $\mathcal{K} \subseteq \mathcal{L}$ *and* $\leq'$ *is a pre-ordering on the non-axiom premises* $\mathcal{K} \setminus K_n$*. Here,* $\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_p \cup \mathcal{K}_a$ *where these subsets of* $\mathcal{K}$ *are disjoint:* $\mathcal{K}_n$ *is the (necessary)* axioms *(which cannot be attacked);* $\mathcal{K}_p$ *is the* ordinary premises *(on which attacks succeed contingent upon preferences), and;* $\mathcal{K}_a$ *is the* assumptions *(on which attacks are always successful).*

We refer the reader to the work of Bondarenko et al. [1997] in relation to the employment of assumptions as premises and their vulnerability against attacks.

Arguments are now defined, where for any argument A, $\texttt{Prem}$ returns all the formulæ of $\mathcal{K}$ (*premises*) used to build A; $\texttt{Conc}$ returns A's conclusion; $\texttt{Sub}$ returns all of A's sub-arguments; and $\texttt{Rules}$ returns all rules in A.

Definition 13 (Argument) *An argument A on the basis of a knowledge base $(\mathcal{K}, \leq')$ in an argumentation system $(\mathcal{L}, ^{-}, \mathcal{R}, \leq)$ is:*

1. *φ if $\varphi \in \mathcal{K}$ with:*
   $\texttt{Prem}(A) = \{\varphi\}$
   $\texttt{Conc}(A) = \varphi$
   $\texttt{Sub}(A) = \{\varphi\}$
   $\texttt{Rules}(A) = \emptyset$

2. *$A_1, \ldots A_n \rightarrow/\Rightarrow \psi$ if $A_1, \ldots, A_n$ are arguments such that there exists a strict/defeasible rule $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow/\Rightarrow \psi$ in $\mathcal{R}_s/\mathcal{R}_d$.*
   $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$
   $\texttt{Conc}(A) = \psi$
   $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$
   $\texttt{Rules}(A) = \texttt{Rules}(A_1) \cup \ldots \cup \texttt{Rules}(A_n) \cup \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow/\Rightarrow \psi\}$

In Modgil and Prakken [2011], the notion of *c-consistent* arguments is also introduced, so that classical logic approaches to argumentation (e.g., Besnard and Hunter [2008]), including those that accommodate preferences (Amgoud and Cayrol [2002a]), can be captured as instances of *ASPIC*$^+$. These approaches require that the premises of arguments are consistent. Hence, Modgil and Prakken [2011]'s c-consistent arguments are defined as those whose premise sets cannot be extended by strict rules to obtain arguments with contradictory conclusions.

Three kinds of *attack* are defined for *ASPIC*$^+$ arguments. B can attack A by attacking a premise or conclusion of A, or an inference step in A. For the latter *undercutting* attacks, it is assumed that applications of inference rules can be expressed in the object language; the precise nature of this naming convention will be left implicit. Some kinds of attack succeed as *defeats* independently of preferences over arguments, whereas others succeed only if the attacked argument is not stronger than the attacking argument. In this sense, attacks can be distinguished as either preference-dependent or preference-independent, where the former's success as defeats is determined by a preference ordering on the constructed arguments.

So as to not compromise the generality of our framework we make no assumptions on the properties of $\leq$. Nevertheless, for the purpose of our work we will utilise the two pre-orderings, on defeasible rules and on the non-axiom premises (we assume their usual strict counterparts, i.e. $l < l'$ iff $l \leq l'$ and $l' \not\leq l$), for defining an ordering $\preceq$ on the constructed arguments. Unlike Prakken [2010] we explicitly define a function p that takes as input a knowledge base in an argumentation system (and so the defined arguments and orderings on rules and premises) and returns an ordering on the constructed arguments. Henceforth, we assume the strict counterpart $\prec$ of $\preceq$. We however note, that the definition of defeat does not rely on these particular pre-orderings.

Definition 14 (Attacks: $\mathcal{C}$) *An argument A attacks an argument B iff A* undercuts, rebuts *or* undermines *B, where:*

- *A* undercuts *argument B (on B') iff $\texttt{Conc}(A) \in \overline{B'}$ for some $B' \in \texttt{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \psi$*
- *A* rebuts *argument B (on B') iff $\texttt{Conc}(A) \in \overline{\varphi}$ for some $B' \in \texttt{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \varphi$. In such a case A* contrary-rebuts *B iff $\texttt{Conc}(A)$ is a contrary of φ*
- *A* undermines *B (on B') iff $\texttt{Conc}(A) \in \overline{\varphi}$ for some $B' = \varphi$, $\varphi \in \texttt{Prem}(B) \backslash \mathcal{K}_n$. In such a case A* contrary-undermines *B iff $\texttt{Conc}(A)$ is a contrary of φ or if $\varphi \in \mathcal{K}_a$*

An undercut, contrary-rebut, or contrary-undermine attack is said to be preference-independent, *otherwise an attack is* preference-dependent.

Definition 15 (Defeats) *An argument A defeats an argument B (denoted $A \rightarrow B$) iff A attacks B (denoted $A \rightharpoonup B$) on B', and either: $A \rightharpoonup B$ is* preference-independent, *or; $A \rightharpoonup B$ is* preference-dependent *and $A \nprec B'$.*

Definition 16 (Argumentation Theory) *An argumentation theory is a triple $AT = (AS, KB, p)$ where AS is an argumentation system, KB is a knowledge base in AS and:*

$$p : AS \times KB \longrightarrow \preceq \tag{2.1}$$

such that $\preceq$ is an ordering on the set of all arguments that can be constructed from KB in AS.

We refer the reader to the work presented by Prakken [2010] for ways in which such a function (p), to which we will be referring as *preference function*, would define $\preceq$ according to the *weakest* or *last link principles*.

The justified arguments under the full range of Dung [1995] semantics can then be defined. To recap:

- A Dung framework consists of a set of arguments $\mathcal{A}$ and a binary relation `attacks` over $\mathcal{A}$.
- $S \subseteq \mathcal{A}$ is *conflict-free* iff $\forall X, Y \in S$, $(X, Y) \notin$ `attacks`. $X \in \mathcal{A}$ is acceptable with respect to some $S \subseteq \mathcal{A}$ iff $\forall Y$ such that $(Y, X) \in$ `attacks` implies $\exists Z \in S$ such that $(Z, Y) \in$ `attacks`.
- A conflict-free set S is an *admissible* extension iff $X \in S$ implies X is acceptable with respect to S.
- A conflict-free set S is *complete* extension iff $X \in S$ iff X is acceptable with respect to S.
- A conflict-free set S is preferred extension iff it is a set inclusion maximal complete extension.
- A conflict-free set S is *the* grounded extension iff it is the set inclusion minimal complete extension.
- For $s \in$ {complete, preferred, grounded}, X is *sceptically* or *credulously* justified under the s semantics if X belongs to all, respectively at least one, s extension.

Thus, if $\mathcal{A}$ is the set of (c-consistent) arguments on the basis of an *ASPIC*$^+$ argumentation theory AT, $\mathcal{C}$ the attack relation over these arguments, and $\mathcal{D}$ the defeat relation obtained from $\mathcal{C}$ and the preference ordering $\preceq$, then letting $\mathcal{D}$ be the binary relation `attacks`, the justified arguments of AT are the justified arguments of the Dung framework $(\mathcal{A}, \mathcal{D})$. Prakken [2010] shows that under some intuitive assumptions on the strict knowledge and the preference relation $\preceq$, *ASPIC*$^+$ satisfies all of Caminada's [2007] rationality postulates for argumentation.

Modgil and Prakken [2011] argue that unlike Prakken [2010], and other works that derive defeat relations from attack relations, conflict free-ness should be defined with respect to attacks ($\mathcal{B} = \mathcal{C}$ in the above definition of conflict-free), and the defeat relation should only be used to define the acceptability of arguments ($\mathcal{B} = \mathcal{D}$ in the above definition of acceptability). They then show that under the above assumptions on strict knowledge and preferences, the key results for Dung's theory are preserved, and Caminada's [2007] rationality postulates are satisfied. Henceforth we will assume evaluation of justified arguments as defined by Modgil and Prakken [2011].

In summary Prakken [2010] and subsequently Modgil and Prakken's [2011] *ASPIC*$^+$ provides a general framework that accommodates a number of possible logical approaches to argumentation[1] and satisfies Caminada's [2007] rationality postulates.

2.4 Agent Dialogues

While argument games may resemble the form of a two-person debate or discussion, one needs to keep in mind that these proof procedures are actually monologues where an agent is actually opposing itself, by simulating a dialogue between two imaginary sides. An interesting question to explore is: "*How can argument games be generalised to actual two-person dialogues?*". Interest in responding to this question mainly arises from the fact that in dialogues knowledge is distributed amongst the participants. Thus properties that characterise a participant's arguments within a framework instantiated from its own knowledge-base

[1] E.g. Modgil and Prakken [2011] and Prakken [2010] show that assumption based (Bondarenko et al. [1997]) and classical logic argumentation are instances of *ASPIC*$^+$.

(e.g. extension membership), may not hold in a framework incrementally built through sequential introduction of arguments by both sides in a dialogue. In addition, it is not possible to impose to the participants of a certain kind of a dialogue game to comply to their dialogical roles—for example in the case of a persuasion dialogue, to convince their opponents of their own view—or to be truthful. Thus, it becomes essential that criteria under which an argument can be deemed acceptable with respect to information collected through a dialogue are formally defined, so as to account for these differences.

However, prior to responding to this question one needs to account for the different types of dialogues studied in the literature as well as for the necessary components for the development of dialogue frameworks. For doing so we provide a general overview on dialogue games, based on the work of McBurney and Parsons [2009].

2.4.1 Types of Dialogues

Relying on the work of Walton and Krabbe [1995], "a model of human dialogues", McBurney and Parsons [2009] present a summarise the dialogue categories in six different types, based on the different objectives that characterise them.

1. **Information-Seeking Dialogues:** One of the participants seeks the answer to some question(s) from the other based on the belief of the first that the latter has it.

2. **Inquiry Dialogues:** Both participants engage in a dialogue in order try to jointly answer a question whose answer is not known to both.

3. **Persuasion Dialogues:** A participant tries to convince another to accept a proposition that the last does not currently endorse.

4. **Negotiation Dialogues:** Both participants bargain over the division of some scarce resource.

5. **Deliberation Dialogues:** The participants collaborate in order to decide the action or the course of actions that they should adopt in order to bring

about some goal.

6. **Eristic Dialogues:** The participants quarrel verbally in order to vent perceived grievances.

We should also mention here a class of dialogues not mentioned in the Walton and Krabbe [1995] taxonomy, referred to as *examination dialogues* (Dunne et al. [2005]; Walton [2006]). The concept of examination dialogue is one in which one participant (the Examiner) questions the other (the Examinee) not in order to discover new information (as is the case in inquiry dialogues), or to make a persuasive case, but rather to asses the extent to which information known to the former is understood by the latter. Here, dependent on the exact context, the Examinee may cooperate (e.g. viva voce environment) or try to evade revealing information (e.g. Detective/Suspect interrogation). Thus in examination dialogues the goal of one player may vary.

It must be noted, that even though dialogues are differentiated based on the objectives they 'impose' to their participants, as McBurney and Parsons [2009] explain, it makes little sense to talk about the goals of a dialogue since the ones who actually have goals are the participants. As a result, it is reasonable to expect that a dialogue is more likely to be a combination of these different types, instead of a specific one. For example, a participant may enter a dialogue in order to negotiate over some resources, while the other could simply be interested in stalling the first, in which case we would have a combination of a negotiation and an eristic dialogue. A similar case is discussed in the work of Gabbay and Woods [2001], which reviews the so-called notion of "stone-walling[1]" by an agent in a persuasion dialogue.

Additionally, McBurney and Parsons [2009] explain that a dialogue can also have phases, and thus transitions between them. For example, participants may enter in a deliberation or in a negotiation dialogue, and change to a persuasion dialogue in the face of some conflict (such an example is given in Section 2.4.4).

Furthermore, and especially in relation to persuasion dialogues which are found at the core of this thesis, we note that it is not always the case that

[1] A stance by which an agent attempts to delay or obstruct (a request, process, or person) by refusing to answer questions or by being evasive.

such dialogues are conducted in an impartial, neutral environment in which both players wish only to objectively reach an agreement on some question at issue. There are a number of contexts in which this view is unrealistic. For example, one may have so-called “hidden agenda” motives (Dunne [2006]; Silverman [2005]) underpinning why for example a proponent wishes to convince its opponent of some proposition p, while at the same time the proponent is uninterested in the truth of p. Furthermore, as Dunne [2003] explains in his work, participants may even act in a ‘non-cooperative’ way (possibly even malicious), contesting dialogue policies or decisions accepted by others, so as to improve some notional individual utility. Such issues bear considerable impact in modelling opponent knowledge and consequently in anticipating one’s strategy and thus need to be taken into account. As it will be discussed in Chapter 3, we account for these possibilities through explicitly modelling a participant’s individual objectives.

2.4.2 Syntax

The form of the utterances as well as well as the rules concerned with the order in which utterances can be introduced into dialogues are issues covered by the syntax of a dialogue game. In order to provide a syntax for these types of dialogues, McBurney and Parsons [2009] present a generic framework adapted from McBurney and Parsons [2002]. In this framework they identify seven basic components:

1. **Commencement Rules:** These define when a dialogue can begin.

2. **Locutions:** These define the legal utterances. For example: *questions*, *propositions*, *contest of an assertion* or *request of justification*.

3. **Rules for combination of locutions:** These define the dialogical contexts under which particular locutions are permitted and/or are obligatory or not.

4. **Commitments:** These rules define the circumstances under which participants incur dialogical commitments by their utterances altering the contents of the participants associated commitment stores.

5. **Rules for combination of commitments:** These define how commitments are combined or manipulated when conflicting or complementary utterances occur.

6. **Rules for speaker order:** These define the order which the speakers can make utterances.

7. **Termination rules:** Rules that define when the dialogue ends.

This set of rules represents the backbone of any dialectical interaction protocol. However, as McBurney and Parsons [2009] explain, one may use different interpretations of them. For example, the notion of *commitments* can have many meanings; it can refer to commitment to consistency, where each part is obliged to have a consistent set of arguments, or commitment to execution where an agent must execute an action or maintain a course of action, or even as the obligation to publicly express your beliefs. It may even refer to commitment to the topic in discussion as is the case in Bench-Capon et al. [2008] concerned with examination dialogues, where the objective is to design dialogues between an interviewer and an interviewee must not have the possibility to evade the issue in concern. In this thesis we are more concerned with commitment to consistency with respect to the utterances employed by a participant in a dialogue.

We note that prior to McBurney and Parsons [2002], there has been work in formulating games using executable specification based on computational logic by Stathis [2000]. All the components discussed in by McBurney and Parsons [2002] are also discussed in this line of work, which however abstracts away from commitments; since these can be formulated as domain specific knowledge.

2.4.3 Dialogue Game Semantics

Semantics concerned with dialogue games provide a shared understanding, to both participants, of the meaning of a single utterance or of a combination of them, and consequently they define the meaning of a dialogue. This shared understanding is also provided to the designers of dialogue protocols as a context for studying and comparing the properties of individual protocols, as well as implementing them, while ensuring that their implementation, in open, distributed agent systems, will be undertaken uniformly.

Drawing from the work of Eijk [2000], McBurney and Parsons [2009] distinguish between different types of semantics for dialogue protocols, as follows:

- **Axiomatic Semantics:** Defines each locution of a communication language in terms of the pre-conditions that must be present before a locution can be uttered, and the postconditions which apply following its utterance. They can be further distinguished as *public* and *private* dictating that all respectively at least some of the preconditions and postconditions which describe states or conditions of the dialogue can be observed by both respectively just one participants/participant.

- **Operational Semantics:** Locutions in combination with agent decision-mechanisms are seen as transition operators, operating successively on the states of some abstract machine, and are used to identify which dialogue states are reachable and which are not.

- **Denotational Semantics:** Each element of the language syntax is assigned a relationship to an abstract mathematical entity (its denotation). The purpose is to derive a semantic mapping of a compound statement in the language from the semantic mapping of its elements; a process referred to as *compositionality*.

In relation to the third type of semantics McBurney and Parsons [2009] refer to the *possible worlds semantics* provided by Kripke [1959] in an attempt to illustrate its application. Nevertheless, as they explain the presence of compositionality is not always guaranteed[1], and thus they draw on *game-theoretic semantics*. Similarly, in these semantics each statement in the language is associated with a conceptual game between two players, and the statement can be deemed true if, for example, a winning strategy[2] exists for the proponent in that game. Research in this thesis orients around this last type of semantics.

[1] A compound statement can have infinite combinations.

[2] The notion of the *winning strategy* is used here in similar sense to how it is used by Modgil and Caminada [2009]; Vreeswijk and Prakken [2000].

2.4.4 Generalising Argument Games to Dialogue Games

Dialogue games can be understood as generalisations of argument games in two ways. The first concerns the fact that in contrast to argument games which are concerned with a single objective—that of proving/disproving the acceptability of an argument currently in dispute—dialogues may have many different objectives. For example, dialogues can be cooperative, e.g. deliberation dialogues, or competitive, e.g. persuasion dialogues. The second concerns how in both argument games as well as in dialogues, a knowledge-base (Δ) is implicitly and incrementally constructed through the exchange of utterances. However, though in argument games these utterances concern the introduction of complete arguments, in dialogues utterances are more expressive and are able to encapsulate the logical constituents of arguments instead of complete arguments. Therefore, as well as Δ, arguments in dialogue games are also implicitly constructed through the exchange of these utterances.

Assume a case of a persuasion dialogue, where you are a lawyer and you have to support that your client is innocent of a crime. You could argue for example that:

- *Argue*: The defended is innocent (i), since he was not present at the scene of the crime ($\neg p$) and not being at scene of the crime implies his innocence ($\neg p \Rightarrow i$).

In a dialogue however, instead of uttering a complete argument, you could start your position simply by claiming:

- *Claim*: i.

which could be countered by:

- *Why*: i?

to which you can then respond with:

- *Argue: i since* $\neg p$, and $\neg p \Rightarrow i$

which can then be *Challenged* and so on.

In this sense, a persuasion dialogue can be seen as more expressive form of an argument game. However, not all dialogues are persuasion dialogues, so why should someone focus on persuasion dialogues alone? The motivation behind focussing on persuasion dialogues derives from the fact that persuasion can be embedded in all other kinds of dialogues, through the integration of utterances like *"why"* or *"challenge"*. Therefore, results from researching persuasion dialogues can impact the study of other kinds of dialogues.

Assume for example a negotiation dialogue where the two participants negotiate over the price of a car. Further assume that the seller's maximum and minimum prices are respectively:

$$\texttt{max}(Seller) = \pounds 30,000 \;\; and \;\; \texttt{min}(Seller) = \pounds 25,000$$

while the buyer's are:

$$\texttt{max}(Buyer) = \pounds 20,000 \;\; and \;\; \texttt{min}(Buyer) = \pounds 15,000$$

Since the maximum price threshold of the buyer is lower than the minimum price threshold of the seller it is evident that the two participants will not reach an agreement. It seems more reasonable for them to engage in a persuasion dialogue in an effort to first define an agreeable range of prices in which they could negotiate.

2.4.4.1 Motivation

Under the scope of denotational/game-theoretic semantics and provided a framework able to support this generalisation, one may then research and identify the required properties of a dialogue protocol, necessary to guarantee the production of *sound* and *complete* dialogues with respect to notions of acceptability used in argument games. In essence, this process concerns proving a correspondence between the two kinds of games with respect to the results they produce.

In simple words, let Δ be the incrementally constructed knowledge that resulted from a dialogue game for an argument A and for a semantics E, then the *soundness* of this dialogue game can be justified if A is deemed acceptable through an argument game with arguments instantiated from that Δ.

As we have already noted at the beginning of Section 2.4, the challenge in guaranteeing this soundness lies in the fact that knowledge in dialogues in contrast to argument games is actually distributed. This means that imposing on participants that they should or should not introduce a legal possible move in the game is at least challenging, since from the outset, awareness of this possibility becomes available only after a move is introduced into the game and not prior to its introduction. This is one of the issues we investigate in this thesis.

2.5 Summary

In this chapter, we presented Dung's [1995] work on developing a highly abstract but simple theory of argumentation, having as central notions those of attack and acceptability. The first notion is understood as a binary relation between arguments in an AF, while the second is expressed through a variety of acceptability semantics that he defines through his framework. Since these acceptability semantics resemble those developed for non-monotonic logics, Dung's work consists a proof of the claim that most major approaches to non-monotonic reasoning in artificial intelligence (AI) and logic programming are special forms of his theory of argumentation.

Furthermore, focussing on the work of Modgil and Caminada [2009] we presented two procedures for testing the acceptability of an argument with respect to a credulous and a grounded semantics, referred to as argument games. The work of these researchers is based on the natural expression of argumentation through dialogue games, though these argument games are actually monologues between two imaginary participants.

We briefly presented Prakken's [2010] work on the development of an abstract framework for argumentation with structured arguments, referred to as $ASPIC^{+}$. $ASPIC^{+}$ distinguishes between different kinds of argument premises, while it accommodates preferences that characterise the logical constituents of arguments from which a defeat relation between arguments can be derived. In addition it captures a number of well known logical approaches to argumentation and satisfies Caminada's [2007] rationality postulates.

Finally, we reviewed the work of McBurney and Parsons [2009] who focus on dialogue game protocols for agent interaction and argument, through analysing these protocols from both a syntactical as well as a semantical perspective. Through their work they provide an analytic framework for the development of different kinds of dialectical protocols. The development of one such framework is one of the concerns of this thesis, specifically related to argumentative systems for persuasion dialogues which is presented in Chapter 3.

Chapter 3

Strategising in Agent Dialogues

In recent years there has been much interest in researching the notion of strategy within argumentation based dialogues.That is, given a choice of locutions to put forth during the course of a dialogue, which should an agent choose, and under what circumstances. Most approaches either concern the heuristics based on which a strategy may be implemented in a dialogue game, or the analysis of the strategic aspects that characterise argument interactions based on a game theoretic perspective. However, much of the work in this field assumes an abstract form of arguments, which fails to take into account the logical content and structure of arguments.

In this chapter we provide a general framework for dialogue games in Section 3.1, in which rational agents can strategise, based on their own beliefs, and their assumptions about their interlocutor's beliefs, environment and goals. We then formalise a specific persuasion dialogue as an instance of the framework, and show how agents can strategise. Since strategic considerations require that the content and structure of arguments is explicitly modelled, we build on the recent $ASPIC^+$ model of argumentation presented in Chapter 2. The latter explicitly models the logical content and structure of arguments, while at the same time accommodating many existing logics for argumentation allowing us to not compromise the generality of our framework. Relying on this general framework we instantiate a persuasion dialogue instance of it that we present in section 3.2. In contrast with existing work on persuasion dialogues (e.g., Prakken [2006]; Riveret et al. [2007, 2008]), we allow the participants to not only move arguments that *attack* those of their interlocutor, but also (possibly contradictory) *preferences* that undermine the success of these attacks as *defeats*.

We define two persuasion dialogue protocols that conform to the grounded and credulous (admissible/preferred) semantics, and prove *soundness* and *fairness* (a form of completeness) results for these dialogues in Section 3.3.2. Finally, Section 3.4 concerns an investigation of strategic considerations in persuasion dialogue instances of our framework. We show how participants may strategise based on their beliefs about their interlocutor's knowledge, and how such considerations need to account for the logical content of arguments. Particularly, we focus on showing how the abstract approach fails to accommodate the dynamics of dialogue, whereby new arguments may be constructed during the course of the dialogue process.

3.1 A General System for Dialogue

In this section we focus on providing a general system for dialogue, based on $ASPIC^+$, and adapting the system proposed in the ASPIC project, Amgoud et al. [2006]. In general, a dialogue system usually consists of three basic elements. Namely, these are: the reasoning model employed by the participating agents; the set of rules responsible for regulating a dialogue game and; the basic elements that allow for the implementation and employment of strategies by the participants.

The first characterises an agent's reasoning, necessary for inferring and justifying conclusions, explaining facts and making decisions. With respect to dialogues, decision making concerns defining one's objectives in relation to its main goals; choosing the appropriate type of dialogue to initiate, for resolving a certain problem and; choosing the content of a move at a given step. Then, given a certain dialogue type, we rely on different sets of rules—protocols—for regulating the course of a dialogue game. That is, for defining the participants' turn-taking; the set of possible moves allowed to be deployed in a participant's turn and; the game's termination rules. Finally, it is also necessary that we provide the means through which a participant will be able to implement and employ a strategy, so as to increase its chances of satisfying its self-interested objectives. These concern the heuristics and information upon which a strategy can be based.

3.1.1 Agent Theories

We begin by assuming an environment of multiple agents $Ag_1, \ldots, Ag_\nu$, where each Ag_i can engage in dialogues in which its strategic selection of locutions may be based on what Ag_i believes its interlocutor (in the set $Ag_{j \neq i}$) knows[1]. Accordingly, and in similar sense to the approach employed by Oren and Norman [2010], each Ag_i maintains a model of its possible opponent agents, though in contrast with Oren and Norman [2010], the model consists of the goals and knowledge other agents may use to construct arguments and preferences, rather than just the abstract arguments and their relations. We assume that all agents share the same contrary relation, the same language $\mathcal{L}$, and the same *way* of defining preferences over arguments (not that they *necessarily* share the same preferences) based on the pre-orderings over non-axiom premises and defeasible rules, i.e. all agents share the same function p.

Definition 17 (Agent Theory) *Let $Ags = \{Ag_1, \ldots, Ag_\nu\}$ be a set of agents. For $i = 1 \ldots \nu$, the theory of Ag_i is a tuple $AgT_i =< S_{(i,1)}, \ldots, S_{(i,\nu)} >$ such that or $j = 1 \ldots \nu$, each sub-theory $S_{(i,j)} = \langle AT_{(i,j)}, \mathcal{G}_{(i,j)} \rangle$ where $AT_{(i,j)}$ is what Ag_i believes is the argumentation theory $(AS_{(i,j)}, KB_{(i,j)}, p_{(i,j)})$ of Ag_j and $\mathcal{G}_{(i,j)}$ is what Ag_i believes are the goals of Ag_j, and:*

- *in the case that $j = i$, $AT_{(i,j)}$ and $\mathcal{G}_{(i,j)}$ are respectively Ag_i's own argumentation theory and goals.*
- *for $i, j, k, m = 1 \ldots n$, let $S_{(i,j)}$, $S_{(k,m)}$ be any two distinct sub-theories of the form $\langle AT_{(i,j)}, \mathcal{G}_{(i,j)} \rangle$, $\langle AT_{(k,m)}, \mathcal{G}_{(k,m)} \rangle$, where $AT_{(i,j)} = (AS_{(i,j)}, KB_{(i,j)}, p_{(i,j)})$, $AT_{(k,m)} = (AS_{(k,m)}, KB_{(k,m)}, p_{(k,m)})$. We then assume that:*
  - $p_{(i,j)} = p_{(k,m)}$
  - *$\mathcal{L}_{(i,j)} = \mathcal{L}_{(k,m)}$, $^-(i,j) = {}^-(k,m)$, where $\mathcal{L}_{(i,j)}$ ($^-(i,j)$) and $\mathcal{L}_{(k,m)}$ ($^-(k,m)$) are the languages (contrary relations) in $S_{(i,j)}$ and $S_{(k,m)}$ respectively.*

[1] We clarify that "believe" and "know" should be understood informally here, and are not related to some modal epistemic logic.

		1	2	.	i	.	ν
		Ag_1	Ag_2	$\dots$	Ag_i	$\dots$	Ag_ν
1	AgT_1	$S_{(1,1)}$	$S_{(1,2)}$	$\dots$	$S_{(1,i)}$	$\dots$	$S_{(1,\nu)}$
2	AgT_2	$S_{(2,1)}$	$S_{(2,2)}$	$\dots$	$S_{(2,i)}$	$\dots$	$S_{(2,\nu)}$
.	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
i	AgT_i	$\dots$	$\dots$	$\dots$	$S_{(i,i)}$	$\dots$	$\dots$
.	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
ν	AgT_ν	$S_{(\nu,1)}$	$S_{(\nu,2)}$	$\dots$	$S_{(\nu,i)}$	$\dots$	$S_{(\nu,\nu)}$

(a)

		1	2	3	4	5
		$\mathcal{K}$	$\leq'$	$\mathcal{R}$	$\leq$	$\mathcal{G}$
1	$S_{(i,1)}$	$\mathcal{K}_{(i,1)}$	$\leq'_{(i,1)}$	$\mathcal{R}_{(i,1)}$	$\leq_{(i,1)}$	$\mathcal{G}_{(i,1)}$
2	$S_{(i,2)}$	$\mathcal{K}_{(i,2)}$	$\leq'_{(i,2)}$	$\mathcal{R}_{(i,2)}$	$\leq_{(i,2)}$	$\mathcal{G}_{(i,2)}$
.	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
i	$S_{(i,i)}$	$\mathcal{K}_{(i,i)}$	$\leq'_{(i,i)}$	$\mathcal{R}_{(i,i)}$	$\leq_{(i,i)}$	$\mathcal{G}_{(i,i)}$
.	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
ν	$S_{(i,\nu)}$	$\mathcal{K}_{(i,\nu)}$	$\leq'_{(i,\nu)}$	$\mathcal{R}_{(i,\nu)}$	$\leq_{(i,\nu)}$	$\mathcal{G}_{i_\nu}$

(b)

Table 3.1: a) A matrix $\mathcal{M}_{\nu,\nu}$, where each row i represents a distinct agent theory, b) the distinct sets of logical elements found in each sub-theory of a AgT_i

One can understand a tuple $S_{(i,j)}$ as the formal representation of an OM, which we illustrate for presentation convenience as a row in a two dimensional matrix that represents one's agent theory AgT_i, as depicted in Table 3.1b. The set elements of a sub-theory represent a modeller's assumptions about the beliefs of its opponent, whether these are, inference rules ($\mathcal{R}$), premises ($\mathcal{K}$) or preferences. This is also the case with the modelled goals ($\mathcal{G}$) of an opponent, which are not open and as with everything else are also subject to belief.

One may represent the set of all the distinct agent-theories of a number of agents equal to ν operating in a multi-agent environment, through a two dimensional matrix $\mathcal{M}_{\nu,\nu}$ of the form presented in Table 3.1a (note that henceforth we may omit subscripts identifying pre-orderings and rules specific to a given agent). We refer to this matrix as Multi-Agent Omni-base (MAOB).

Since we are dealing with the concept of strategising in dialogues, it makes sense that the effectiveness of the proposed approach is evaluated with respect to the level of correspondence of the assumed information (the OM), in relation to the beliefs of a participant's interlocutor. In this sense, the cases where an agent's (Ag_i's) beliefs about its interlocutor's beliefs ($Ag_{j\neq i}$'s) match exactly the latter's actual beliefs and vice-versa, appear in the cases where $S_{(i,j)} = S_{(j,j)}$ and $S_{(j,i)} = S_{(i,i)}$ respectively. These cases explicitly represent the *optimal* opponent-modelling cases for each of the participating agents, where their assumptions about the beliefs of their interlocutors are both valid and encapsulate all the information know to an opponent.

In general, we differentiate between the following possible cases for $i \neq j$:

- `False`: iff $S_{(i,j)} \cap S_{(j,j)} = \varnothing$
- `Partial`: iff $S_{(i,j)} \cap S_{(j,j)} \neq \varnothing$
- `Contained`: iff $S_{(i,j)} \subset S_{(j,j)}$
- `Identical`: iff $S_{(i,j)} = S_{(j,j)}$
- `Excessive`: iff $S_{(i,j)} \supset S_{(j,j)}$.

These typifications may also be used to characterise the level of correspondence of each of the assumed logical set elements $\mathcal{K}, \leq', \mathcal{R}, \leq, \mathcal{G}$ of a sub-theory $S_{(i,j)}$, with the elements representing the actual beliefs of an agent's interlocutor $S_{(j,j)}$, so as to provide a more thorough description.

Furthermore, we should note that we do not assume any of the logical elements of the components of each sub-theory $S \setminus \{\mathcal{K}_n, \mathcal{R}_s\}$ of an AgT_i to be necessarily consistent. For example, it may be the case that both $\neg q$ and q are elements of $\mathcal{K}_p$. Nonetheless, we do assume that the elements of both $\mathcal{K}_n$ and $\mathcal{R}_s$ are consistent, since those refer to an agent's axiomatic beliefs. However, the latter assumption does not prevent the inference of inconsistent conclusions. For example, it may very well be that an agent believes the axiom premises p and s, while also believing the defeasible inference rules $p \Rightarrow q$ and $s \Rightarrow \neg q$.

3.1.2 Contexts

It is often the case that a dialogue takes place in a *context* with reference, for example, to a legal system, a set of conformities, a domain or maybe a system description. Formally, that is to engage in a dialogue game where both participants share a set of dogmatic beliefs. In these cases one may assume that the logical elements describing these dogmatic beliefs are integrated into the participating agents' own sub-theories, as elements of their axiomatic beliefs ($\mathcal{K}_n$ and $\mathcal{R}_s$). In this sense, we define a context as follows:

Definition 18 *A context $\mathscr{C} \subseteq \mathcal{L}$ is a tuple $< \mathcal{K}_n^{\mathscr{C}}, \mathcal{R}_s^{\mathscr{C}} >$, the elements of which respectively represent both the axiom premises and the strict rules shared by the participating agents $Ag_1, \ldots, Ag_\nu$, such that:*

$$\mathcal{K}_n^{\mathscr{C}} \subseteq K_{n(i,i)} \ \ \textit{and} \ \ \mathcal{R}_s^{\mathscr{C}} \subseteq R_{s(i,i)}, \ \ \textit{for} \ \ i = 1 \ldots \nu.$$

3.1.3 The System

Though there is a number of dialogue types ($\mathcal{DT}$) worth researching through this strategical perceptive, we mainly focus on a particular class of dialogues as the latter is expressed by the Walton and Krabbe [1995] dialogue typology. According to this typology different dialogues can be distinguished based on the commitments, the type of starting point, and the type of dialogical goal that characterises them. Namely, these different types are: $\mathcal{DT}$ = {`Persuasion, Negotiation, Inquiry, Deliberation, Information-seeking, Eristics`}. Out of these six classes we will be accounting for the first five, given that `'Eristic'` dialogues represent a formalisation for dialectical quarrelling, which suggests that rational analysis and strategy determination is problematic for them.

As Austin [1962] explains, one can capture these different dialogue types through a set of speech-acts ($\mathcal{SA}$), which are assumed to cover the full range of dialectical interactions found in different dialogue types. Examples of speech acts commonly employed in dialogues include `Accept`, `Argue`, `Challenge`, `Disinform`, `Inform`, `Offer`, `Question`, `Reject`, and `Request`.

In a similar sense to Amgoud et al. [2006], we also assume that locutions exchanged between participants in dialogues are conjoinings of speech-acts augmented with content. In the proposed approach, we define content with respect to the elements of an $ASPIC^+$ argumentation theory and its defined arguments and preferences. Of course, for any given dialogue type, not all combinations of speech-acts and content are valid.

In the following definition which describes the form of locutions used in our research context, we also define a function that validates combinations for given dialogue types.

Definition 19 (Locution) *Let $\mathcal{SA}$ be a set of speech-acts and AT the argumentation theories of an agent Ag. Let $\mathcal{A}$ and $\preceq$ be the arguments and preferences defined by AT as in Section 2.3.1. Then:*

- *a locution $\mathcal{M}$ is of the form $f : x$, where:*
  - $f \in \mathcal{SA}$
  - $x \in \mathcal{DT}$ *or* $x \in \texttt{Content}$ *where* $\texttt{Content} \in \{\mathcal{L}, \leq', \leq, \preceq, \mathcal{R}, \mathcal{A}\}$
- *we may write* $\texttt{Act}(\mathcal{M}) = f$ *to denote the speech-act f and* $\texttt{Content}(\mathcal{M}) = x$ *to denote the content x*
- *the set of all possible locutions which may be extracted from an AT is expressed as M^{all}*
- *for each $t \in \mathcal{DT}$, let $V^t : \mathcal{M} \longrightarrow \{Success, Failure\}$. Then, a locution $\mathcal{M} \in M^{all}$ is valid iff:*
$$V^t(\mathcal{M}) = Success.$$
  All valid locutions for a specified $t \in \mathcal{DT}$ is denoted as M_t.

A dialogue is then defined as a sequence of *dialogue moves*. Borrowing again from Amgoud et al. [2006], we define a dialogue move $\mathcal{DM}$ as a 4-tuple as follows:

Definition 20 (Dialogue Move) *For a dialogue type $t \in \mathcal{DT}$, a dialogue move $\mathcal{DM}$ is a tuple $< Ag_i, Ag_j, \Sigma\mathcal{M}', \Sigma\mathcal{M} >$, such that:*

- *Ag_i and Ag_j are the interlocutors where $i \neq j$, while we say that:*
  - $\texttt{Speaker}(\mathcal{DM}) = Ag_i$
  - $\texttt{Hearer}(\mathcal{DM}) = Ag_j$
- $\texttt{Locution}$*($\mathcal{DM}$)* $= \Sigma\mathcal{M} \subseteq M^t$ *where:*
  - $\Sigma\mathcal{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_\mu\}$
  - *where μ is the number of all* ***legal*** *locutions allowed in a given turn*

- $\texttt{Target}(\mathcal{DM}) = \varSigma\mathcal{M}'$ *where:*
  - $\varSigma\mathcal{M}' = \{\tau\mathcal{M}'_1, \ldots, \tau\mathcal{M}'_\mu\}$ *such that each* $\tau\mathcal{M}'_k$*, is a* ***set*** *of the target locutions of the corresponding* $\mathcal{M}_k \in \varSigma\mathcal{M}$*, for* $1 \leq k \leq \mu$*.*

The set of all possible dialogue moves that may be constructed based on a set M^t*, is denoted as* DM_t^{all}*.*

It is worth noting that the sets of target locutions in $\varSigma\mathcal{M}'$ might not be disjoint, since there might exist more than one locution in $\varSigma\mathcal{M}$ with the same target locution.

Definition 21 (Dialogue) *A dialogue* $\mathscr{D}$ *is a finite sequence* $< \mathcal{DM}_0, \ldots, \mathcal{DM}_n >$ *where:*

- $\mathcal{DM}_0, \ldots, \mathcal{DM}_n \in DM_t^{all}$
- $\texttt{Target}(\mathcal{DM}_0) = \theta$ *where* θ *is a locution*
- $\forall i \geq 1, \exists\ j < i$ *such that:*
  - $\texttt{Target}(\mathcal{DM}_i) = \varSigma\mathcal{M}'$*, where* $\varSigma\mathcal{M}' = \texttt{Locution}(\mathcal{DM}_j)$*, and*
  - $\texttt{Speaker}(\mathcal{DM}_i) = \texttt{Hearer}(\mathcal{DM}_j)$
- *any sub-sequence* $< \mathcal{DM}_0, \ldots, \mathcal{DM}_{k<n} >$ *is denoted as* d_k*, while the extension of* d_k *with a dialogue move* $\mathcal{DM}_i$ *will be denoted by the juxtaposition:*

$$d_k.\mathcal{DM}_i$$

- *The set of all moves employed in a* $\mathscr{D}$ *is denoted as:*

$$\Sigma\mathscr{D} = \{\mathcal{DM}_0, \ldots, \mathcal{DM}_n\}.$$

3.1.4 The Protocol

Every dialogue game is characterised by a protocol. Its purpose is to define a dialogue in three aspects. The first concerns the participants' *turn-taking*, which defines how the participating agents may exchange turns in a game, as well as the number of locutions that a participant can assert in a single turn. The second concerns the *legality* of a locution. It basically refers to the set of rules based on which the set of the available choices that can serve as response options for an agent in its next turn is defined.

This set of rules is responsible for regulating the participants' *consistency*, as well as the *dialogical coherence* of a game as well (Prakken [2006]). The first concerns the application of restrictions with respect to the introduction of locutions that compromise the consistency of a speaker's commitments, in relation to previously introduced locutions. This aspect is can be regulated by means of a *legal-move* function as is the case in the work of Jakobovits and Vermeir [1999] where a self-consistency dialogue type is introduced in which the set of arguments uttered by any player has to be consistent.

Dialogical coherence concerns whether a move is allowed to serve as a reply, and it is defined either only in relation to the last move asserted in the game, or with respect to any previously asserted moves (*backtracking*). Lastly, the third aspect concerns the *termination rules* of the game, after the termination of which success or failure for each participant with respect to the dialogue game's result must be defined. This can only be done based on the participants' *roles* in a dialogue. In fact, it is often the case that the legality and turn-taking aspects of a protocol are subjectively defined with respect these roles. Given the impact of a participant's role in defining a protocol, we begin with providing a formal defintion of the notion of roles in dialogue games.

3.1.4.1 Roles

The participants of a dialogue are usually characterised by different roles with respect to the dialogue's type. For instance, in a persuasion dialogue these roles can be those of the *proponent* whose aim is to persuade its interlocutor of the truth of a claim φ, and the *opponent* who tries to disprove φ. In a negotiation dialogue the participating agents can simply be understood as negotiators, sharing the dynamic roles of the *proposer* and the *recipient* during the course of the game. Depending on the type of the dialogue, these roles can either be static, characterising a participant throughout the whole dialogue process, or dynamic, meaning that during the course of the game the roles might change between the participants.

These roles are characterised by a *dialogical goal* which defines the objective of a dialogue. In the case where a dialogical goal complies with an agent's self-interested goals, we can then assume this goal to be integrated into the agent's goals set $\mathcal{G}$. Nonetheless, regardless of whether such compliance applies or not—regardless of whether an agent will choose to commit to its roles 'obligations'—each participant of a dialogue is still characterised by a role, and thus correspondingly by different sets of turn-taking rules and legality restrictions.

A participant is assigned a role in a dialogue game based on the following definition:

Definition 22 (Roles) *Let $P = \{Ag_1, \ldots Ag_p\} \subseteq Ags$ be the participating agents in a dialogue $\mathscr{D}$ characterised by a dialogue type $t \in \mathcal{DT}$ and is thus associated with a set of roles $\mathcal{R} = \{r_1, \ldots, r_k\}$. Then a function* `Roles` *is applied on P, t and $\mathcal{R}$ such that:*

$$\texttt{Roles}(P, t, \mathcal{R}) \longrightarrow r \tag{3.1}$$

3.1.4.2 Commitments

Another important role of the protocol is to regulate a participant's consistency. This, relates to its *commitment store*. Prakken [2006] refers to the expectancies created by commitments in a dialogue game, as dialectical obligations. For instance, in many systems the interlocutor is assumed to be committed to a proposition bearing the burden of supporting it when challenged, while additionally it is further assumed that she cannot go against that proposition later in the dialogue, since this would imply an inconsistency about one's beliefs. McBurney and Parsons [2002] argue that given that dialogical interactions are intended to lead to the achievement of some wider objective, they implicitly define commitments in terms of future actions or propositions external to the dialogue procedure. Thus, they support that one should view commitments as semantic mappings between locutions and subsets of some set of utterances that may express actions or beliefs that are, again, external to the dialogue. However, in a much more simplified approach, Prakken [2006] argues that:

> *"Strictly speaking, the only dialectical obligation that a participant has, is making an allowed move when it is one's turn."*

For the purpose of the proposed model, we assume that the dialogue takes place using *commitment stores* in a similar sense that these are employed by Hamblin [1970] and Mackenzie [1979], the purpose of which is to maintain all statements to which each interlocutor has committed during the course of the game. We assume a distinct commitment store for each participant in the set $\{Ag_1, \ldots, Ag_\nu\}$, which we assume to be empty at the beginning, and their content being constantly updated in each turn according to function U_{CS}, based on the effects of the last dialogue move introduced into the game on the content of the commitment store.

One may then rely on these commitment stores for partially determining the legality of moves, for defining the contents of future dialogue moves, as well as for strategising (as will be discussed in Section 3.4).

Definition 23 (Commitment stores) *Given a set of agents:*

$$P = \{Ag_1, \ldots Ag_p\} \subseteq Ags$$

participating in a dialogue $\mathscr{D} = < \mathcal{DM}_0, \ldots, \mathcal{DM}_n >$, *for any agent* $Ag_i \in P$ *we define the evolution of its* commitment store CSi *as follows:*

1. $CSi_0 = \emptyset$

2. *For* $j = 0 \ldots n$, CSi_{j+1} *is obtained by updating* CSi_j *with the effects of the dialogue move* $\mathcal{DM}_j$, *through the use of a function* **`update`** *such that:*

$$\texttt{update}(CSi_j, \mathcal{DM}_j) \longrightarrow CSi_{j+1}. \tag{3.2}$$

We should clarify that every participant has full knowledge of the commitments stores of its opponents as those evolve while in a dialogue.

3.1.4.3 Turn-taking & Backtracking

As Prakken [2005] explains, many current dialogue systems impose a rather rigid control structure on dialogues with respect to turn-taking. For instance, they are mostly characterised by protocols that are restrictive in relation to the number of legal locutions allowed to be moved into the game in a single turn. Alternatively, other protocols allow for both *single-* and *multi-* locutions to be asserted in the dialogue. In the proposed framework we account for the employment of both kinds of protocols, given that we allow for the content of each dialogue move to be either a single locution ($\mathcal{M}$) or a set of locutions ($\Sigma\mathcal{M}$).

We assume the number of elements $\mu \leq |\Sigma\mathcal{M}|$ for a set $\Sigma\mathcal{M}$ of legal locutions (we discuss legality in Section 3.1.4.4) that are allowed to be introduced in a single turn, to rely on a game's turn-taking rules, imposed through a turn-taking function. Particularly, this function is responsible for defining both μ, as well as the participant to speak next. In single-location protocols it holds that $\mu = 1$ while for multi-locutions protocols $1 \leq \mu \leq |\Sigma\mathcal{M}|$.

In addition, whether the target of an introduced move will be related to the last asserted move in the game or to any of the possible moves asserted by one's interlocutor during a game's course is another issue that needs to be regulated. The latter refers to the notion of *backtracking*. Protocols that allow for backtracking are often referred to as *multi-reply* protocols. If backtracking is allowed in a dialogue game then the number of legal locutions allowed at a turn is also defined in terms of the available replies to previous moves, that might have been postponed for later use, rather than just on the replies that can counter the last move of the dialogue.

We will in the following section show how the definition of a set of protocol rules for specific dialogue games implicitly define both turn-taking and backtracking aspects of a game, as well as the legality of moves. We however define a general turn-taking function, for the sake of completeness, as follows:

Definition 24 (Turn-taking) $P = \{Ag_1, \ldots Ag_p\} \subseteq Ags$ *be the participating agents in a dialogue* $\mathcal{D}$*. A turn-taking function is a function* $\texttt{turn}$*, that takes as* *input the current state of a dialogue d and returns a tuple, such that:*

$$\textit{\textbf{turn}}(\mathscr{D}, P) \longrightarrow < Ag \in P, \mu > . \tag{3.3}$$

A more specific form of this function could also take as input whether backtracking is allowed (in the form of a Boolean value) as well as the result applied by a function that regulates the legality of moves which we define in the next section.

3.1.4.4 Legality

Given a language $\mathcal{L}$, we assume a protocol function $\mathscr{P}$ to be responsible for specifying the legal locutions at each stage of the dialogue. $\mathscr{P}$ can be formally expressed as a function applied on the set of all the valid locutions (M^t) with respect to a dialogue $\mathscr{D}$ in a turn k, that produces a subset of M^t such that $\Sigma\mathcal{M} \subseteq M^t$. As discussed in Section 3.1.4.2, this subset must comply to consistency restrictions with respect to a participants commitment store, as well as to additional property requirements in relation to a dialogical objective.

Given the numerous dialogical objectives that follow from the different dialogue types, we assume the application of this function to be subjective with respect to a set of protocol rules, which we further assume to be described in a narrative way, so as to allow for a high level of expressiveness. An example set of these rules is provided in Section 3.2.

Definition 25 (Protocol Function) *For a dialogue $\mathscr{D}$, and with respect to a set of rules, $\mathscr{P}$ is a function applied in each turn of $\mathscr{D}$, such that:*

$$\mathscr{P}(M^t) \longrightarrow \Sigma\mathcal{M} \subseteq M^t. \tag{3.4}$$

3.1.4.5 Termination rules & Result

We generally assume without committing to this, that a dialogue terminates if for a turn k a participant has no possible response. Formally expressed this is when:

$$\mathscr{P}(M^t_k) \longrightarrow \varnothing.$$

The reason behind only providing a loose termination definition, is the possibility of a dialogue ending at any time by one of the participants simply fleeing the dialogue. In this sense one may wish to allow for termination to be decided by the agents themselves.

Regarding the result of a dialogue, noting the observation by McBurney and Parsons's [2002] that agents have goals and not dialogues, we provide a subjective definition with respect to each participant, so as to define one's success or failure in a dialogue game, in relation to one's goals ($\mathcal{G}$). Assume for instance, a case of a participant whose best interest is to lose a game, then winning it would mean failure with respect to its self-interested objective. In this sense we define a dialogue's result as follows:

Definition 26 (Result) *For an agent Ag in the set of participants:*

$$P = \{Ag_1, \ldots, Ag_p\} \subseteq Ags$$

in a dialogue $\mathscr{D}$, function $\texttt{result}$ *associates a dialogue $\mathscr{D}$ with either Success or Failure, with respect to Ag's set of goals $\mathcal{G}$.*

$$\texttt{result} : \mathscr{D} \times \mathcal{G} \longrightarrow \{Success, Failure\}. \tag{3.5}$$

3.1.5 Strategy Function

We provide a general definition of a strategy function that is used to select a locution from amongst the legal locutions (defined by the protocol) available to a participant Ag at any stage in the course of a dialogue. The function makes use of Ag's beliefs about the knowledge of other participants, as well as their commitment stores.

Definition 27 (Strategy function) *Let $P = \{Ag_1, \ldots, Ag_p\} \subseteq Ags$ be the agents participating in a dialogue $\mathscr{D}$ of type $t \in \mathcal{DT}$, whose current form is expressed as d_{k-1}. Let:*

$$\texttt{turn}(d_{k-1}) \longrightarrow < Ag_i, \mu >$$

such that **Speaker**$(\mathcal{DM}_k)= Ag_i$, *and:*

$$\mathscr{P}(M_k^t) \longrightarrow \mathit{\Sigma}\mathcal{M}_k$$

such that `Locution`$(\mathcal{DM}_k)= \mathit{\Sigma}\mathcal{M}_k$ *are the legal locutions for* $\mathcal{DM}_k$. *Then* $\mathcal{S}tr_i$ *is a function applied on* Ag_i*'s agent theory, the commitment stores of all participants, the current state of the dialogue, and the set of legal locutions, such that:*

$$\mathcal{S}tr_i(AgT_i, CS1 \ldots CSp, d_{k-1}, \mathit{\Sigma}\mathcal{M}_k) \longrightarrow \sigma\mathcal{M} \subseteq \mathit{\Sigma}\mathcal{M}_k. \tag{3.6}$$

3.2 Formalising Persuasion Dialogues

We take a similar approach to Prakken [2006] with the following differences and extensions. Firstly, we consider only `Argue` speech-acts, and leave for future work the implicit construction of arguments[1] through the use of other speech-acts. Secondly, persuasion dialogues in our framework are explicitly linked to the *ASPIC*$^+$ framework. Thirdly, we not only define a game for grounded semantics, but also a game for credulous (preferred) semantics. Furthermore, we also state soundness and fairness results for both games. Fourthly, we accommodate the possibility of conflicting preferences by allowing agents to move arguments that attack rather than defeat, and then separately move possibly conflicting preferences (specifically the premise and rule pre-orderings on which argument preferences are based).

In what follows we assume agents Pr and Op with theories as defined in Definition 17, and use subscripts $_{Pr}$ and $_{Op}$ to identify their argumentation theories and commitment stores. Valid locutions are of the form:

$$\texttt{argue}: \ X$$

where X is an *ASPIC*$^+$ argument or X is a tuple (a, b) where $a \subseteq \leq'$, $b \subseteq \leq$; that is, a tuple of pre-orderings on non-axiom premises and defeasible rules

[1]In Prakken [2006], an argument X = 'p since q and q implies p, and q since r and s implies r', can be constructed by first arguing X' = 'p since q and q implies p' and then in response to 'why q', moving X'' = 'q since r and r implies q'. We assume X is moved in a single locution.

respectively.

Intuitively, a locution with content (a, b) provides the basis for defining a preference over arguments moved earlier in the dialogue, via the shared function[1] p described in Section 2.3.1. Thus, if X has been moved (in $\mathcal{DM}_{i+1}$) as an argument attacking Y (in $\mathcal{DM}_i$), then (a, b) may be moved (in $\mathcal{DM}_{i+2}$) as a reply to X, where (a, b)'s orderings over X and Y's contained elements determine (via p) that $X \prec Y$. Henceforth, we may as an abuse of notation reference content of the form (a, b) in terms of the argument ordering it defines (e.g. $X \prec Y$).

The commitment update function can then be defined so that if the locution's content in a move $\mathcal{DM}$ is an argument A or a tuple (a, b), then the commitment store of the move's speaker is updated with $\texttt{Rules}(A) \cup \texttt{Prem}(A)$, respectively a,b; the commitment stores of all other dialogue participants remain the same. Henceforth we will assume that Pr (Op) can move locutions whose content is obtained from their own argumentation theory AT_{Pr} (AT_{Op}) *and the knowledge in the commitment store* CS_{Op} (CS_{Pr}) *of its interlocutor.*

In relation to the moving of preferences as the content of moves in a dialogues, we note that the proposed approach draws from the work of Bench-Capon et al. [2007] who similarly propose a framework that allows arguing with preferences over *values*. In their work, which orients around *practical reasoning*[2] they support that the acceptability of an argument turns not only on what is true, but also on the values and aspirations of the agent to whom the argument is directed. Since agents have different aspirations there is no right/acceptable answer for all of them, and thus rational disagreement is always possible. Thus, since agents cannot specify the relative priority of their aspirations outside of a particular context, this prioritisation has to be part of the practical reasoning process. Similarly, in the case of our context, since we expect agents to not share the same preferences over the arguments they use in dialogues, we allow these preferences to be defined with respect to each of the participants ability to introduce orderings over the constituents of arguments uttered to support their views.

[1] According to which agents are assumed to share the same way of defining preferences over arguments.

[2] Reasoning about what to do in a given situation.

3.2.1 Protocol Rules for Persuasion Dialogue Games

We define *core* protocol rules for persuasion protocols, conducted according to the *credulous* (preferred) and the *grounded* semantics.

The following set of rules simply define what constitutes a legal dialogue. Note that since Pr and Op share the same contrary relations, there is agreement as to whether a given argument attacks another. Henceforth, we will refer to '$\mathcal{DM}^{con}$' as an abuse of notation for $\texttt{Content}(\texttt{Locution}(\mathcal{DM}))$, and may omit the `Argue` speech-act, referring only to its content. Furthermore, we will refer to, and differentiate between, the interlocutors of a dialogue through $I \in \{Pr, Op\}$, such that $\overline{I} = Pr$ if $I = Op$, while $\overline{I} = Op$ if $I = Pr$.

Definition 28 (Legal Persuasion Dialogues) *A dialogue sequence* $\mathscr{D} = <\mathcal{DM}_0, \ldots, \mathcal{DM}_n >$ *is a legal* persuasion *dialogue if:*

1. *The dialogue begins with Pr proposing an argument X for an initialising claim c_{init}, such that $\texttt{Conc}(X) = c_{init}$ and is said to be 'a persuasion dialogue for c_{init}':*
$$\texttt{Speaker}(\mathcal{DM}_0) = Pr, \mathcal{DM}_0^{con} = \texttt{Argue} : X$$

2. *Pr and Op exchange turns such that:*
$$for \quad i = 0 \ldots n-1, \quad if \quad \texttt{Speaker}(\mathcal{DM}_i) = I \quad then \quad \texttt{Speaker}(\mathcal{DM}_{i+1}) = \overline{I}$$

3. *For $i = 1 \ldots n$, the protocol function $\mathscr{P}_{CP}$ is characterised by the following conditions:*

   3.1. if $\texttt{Speaker}(\mathcal{DM}_i) = Pr$ then:

   a) $\mathcal{DM}_i$ is a reply to some $\mathcal{DM}_j$, $j < i$ (i.e., backtracking is allowed), where either:

   i) $\mathcal{DM}_j^{con} = X$, $\mathcal{DM}_i^{con} = Y$ where Y attacks X, or;

   ii) $\mathcal{DM}_j^{con} = X$, $\mathcal{DM}_i^{con} = X \prec Y$ and $\mathcal{DM}_j$ is a reply to some $\mathcal{DM}_k$, $k < j$ and $\mathcal{DM}_k^{con} = Y$, or;

   iii) $\mathcal{DM}_j^{con} = X \prec Y$, $\mathcal{DM}_i^{con} = Y \prec X$;

b) *else* $\mathcal{DM}_i^{con} = \texttt{Argue} : Y$ *such that* $Y \neq X$ *and* $\texttt{Conc}(Y) = c_{init}$

3.2. *if* $\texttt{Speaker}(\mathcal{DM}_i) = Op$ *then:*

a) $\mathcal{DM}_i$ *is a reply to some* $\mathcal{DM}_j$, $j < i$ *(i.e., backtracking is allowed), where either:*

i) $\mathcal{DM}_j^{con} = X$, $\mathcal{DM}_i^{con} = Y$ *where* Y *attacks* X*, or;*

ii) $\mathcal{DM}_j^{con} = X$, $\mathcal{DM}_i^{con} = X \prec Y$ *and* $\mathcal{DM}_j$ *is a reply to some* $\mathcal{DM}_k$, $k < j$ *and* $\mathcal{DM}_k^{con} = Y$*, or;*

iii) $\mathcal{DM}_j^{con} = X \prec Y$, $\mathcal{DM}_i^{con} = Y \prec X$,

The purpose of these rules is simply to give a basic structure to the dialogue game. Notice that participants are not restricted from repeating either their own or their opponent's move in the game. Thus additional rules must be introduced.

3.2.1.1 The Employment of Preference-orderings

Notice that conditions 3.1.a.ii and 3.2.a.ii in Definition 28 allow the moving of a preference over arguments, to invalidate the success of an attack as a defeat, and 3.1.a.iii and 3.2.a.iii allow the moving of a conflicting preference ordering. In essence the employment of preference orderings as contents of dialogue moves is contingent upon the existence of a preference-depended attack relationship between two arguments X and Y.

We assume that participants are able to introduce pre-orderings against arguments, through either *preference attacking* and *preference-rebut attacking* each other:

Let $\mathcal{A}^I$ and $\mathcal{A}^{\overline{I}}$ be respectively the arguments currently introduced into a dialogue game by each of the participants, we define these attacks as follows:

Definition 29 (Preference attack) *A dialogue move* $\mathcal{DM}_i$ *preference attacks another dialogue move* $\mathcal{DM}_j$*, where* $i \neq j$ *and* $j < i$*, if there exists another dialogue move* $\mathcal{DM}_k$*, where* $k \neq j, i$ *and* $k < j$*, such that:*

$$\mathcal{DM}_k^{con} = X \in \mathcal{A}^I \ \ \textit{and} \ \ \mathcal{DM}_j^{con} = Y \in \mathcal{A}^{\overline{I}}$$

where Y *preference-dependent attacks* X*, if* $\mathcal{DM}_i^{con} = \{X \succ Y\}$*.*

Definition 30 (Preference-rebut attack) *A dialogue move $\mathcal{DM}_i$ preference-rebut attacks another dialogue move $\mathcal{DM}_j$, where $i \neq j$ and $j < i$, if:*

$$\mathcal{DM}_i^{con} = \{X \succ Y\} \;\; \textit{and} \;\; \mathcal{DM}_j^{con} = \{Y \succ X\}$$

We note that since through the exchange of moves participants will likely come across new information, with respect to the logical constituents of the arguments exchanged, it may be necessary that they accordingly update their pre-orderings concerned with $\mathcal{K}$ and $\mathcal{R}$. However, in this thesis we do not formally model the mechanism an agent uses to update its priority ordering over rules and premises. We will assume agents use generic principles to do so, e.g. the well know specificity principle, and the temporal principle (which orders newly acquired knowledge over older knowledge).

3.2.1.2 Dialogue Trees & Forests

In a similar sense to Modgil and Caminada [2009], we also resort to representing dialogues in the form of *dialogue trees* since we allow for backtracking. A persuasion dialogue $\mathscr{D}$ can be represented as a tree with n leaf nodes, consisting of n paths from the root node to each of the leaf nodes, where every child node is a dialogue move introduced as a reply to its parent node. Essentially, each new path results from a backtracking move by Pr or Op. We assume these trees to be implicitly instantiated during the dialogue process.

Definition 31 (Dialogue Tree) *A dialogue tree $\mathcal{T}$ is a tuple of the form:*

$$\mathcal{T} = < \Sigma\mathscr{D}, \mathcal{E} >$$

where the elements of the set $\Sigma\mathscr{D}$ of a dialogue sequence $\mathcal{D}$ appear as nodes, and $\mathcal{E}$ is a binary relationship between pairs of $\Sigma\mathscr{D}$ expressed in the form of directed edges, such that $\mathcal{E} \subseteq \Sigma\mathscr{D} \times \Sigma\mathscr{D}$, where:

Root: *The root move is referred to as $\mathcal{DM}_r$ and:*

$$\mathcal{DM}_r^{con} = X \in \mathcal{A}^{Pr} \;\; \textit{where} \;\; \texttt{Conc}(X) = c_{init} \tag{3.7}$$

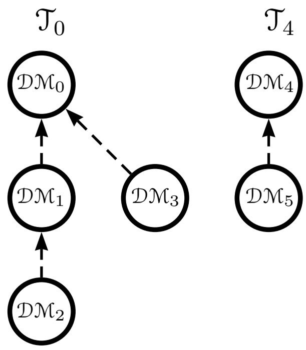


Figure 3.1: A forest of dialogue trees for a dialogue sequence $\mathscr{D} =< \mathcal{DM}_0, \mathcal{DM}_1, \mathcal{DM}_2, D\mathcal{M}_3, \mathcal{DM}_4, \mathcal{DM}_5 >$

Leaf-moves: *There exists a set of leaf-moves* $L_M = \{\mathcal{DM}_1, \mathcal{DM}_2, \ldots, \mathcal{DM}_l, \ldots, \mathcal{DM}_L\}$ *every member of which has no children.*
Father/Child-moves: *We say that a pair of dialogue moves* $(\mathcal{DM}_i, \mathcal{DM}_j)$*, where* $i < j$*, is an element of* $\mathcal{E}$*,* $(\mathcal{DM}_i, \mathcal{DM}_j) \in \mathcal{E}$*, iff there exists an attack relationship between their encapsulated arguments such that* $\mathcal{DM}_j^{con} = Y$ *attacks* $\mathcal{DM}_j^{con} = X \in$*. In this case, the two moves are assumed to be linked in* $\mathcal{T}$ *with a directed edge ending at* $\mathcal{DM}_i$ *and starting at* $\mathcal{DM}_j$*, sharing a father respectively child relationship. Every dialogue move in* $\mathcal{T}$ *can have at most one father-move.*

Another issue concerns the fact that we allow for Pr not only to introduce a legal move into the game that constitutes an attack on one of Op's moves, but also to provide a different argument able to support the initial claim c_{init} of the dialogue in a different way, through the introduction of a different dialogue move that concludes the same thing (Definition 28, 3.1.b). This implies that a single dialogue tree might not be enough to capture a dialogue. Thus it is necessary that we express a dialogue sequence through a number of dialogue trees. This is illustrated in Figure 3.1, where we assume a dialogue $\mathscr{D}$ to be expressed through two dialogue trees $\mathcal{T}_0$ and $\mathcal{T}_4$ indexed in accordance to their corresponding root moves $\mathcal{DM}_0$ and $\mathcal{DM}_4$.

We call the set of these distinct dialogue trees as *dialogue forest* represented as $\mathscr{F}$, while the total number of dialogue trees found in $\mathscr{F}$ is expressed as $f = |\mathscr{F}|$. It is worth noting that a dialogue forest may represent a number of dialogue sequences but not vice-versa.

Given a dialogue tree, one may then refer to the *logically coherent* sub-sequences expressed by the tree's distinct paths as *disputes*. In our work logical coherence is understood as a concept concerned with the structure of a dialogue, in a similar sense to how this notion is used by Prakken [2005]. As Prakken explains, dialogue systems can vary in their structural properties through imposing strict or lenient rules in relation to the possible replies that may follow after a move. For example, imposing that every move must target the opponent's last move in the game, or that it can target any other opponent move in the game. One can characterise a dialogue as *coherent* in the first case, where a line of reasoning[1] can be followed. Similarly in our case, each new dispute that results from a backtracking move by either of the participants satisfies this property.

Definition 32 (Disputes) *For a dialogue $\mathscr{D}$, a dispute $d_{(r \to l)}$ is a sub-sequence of $\mathscr{D}$ of logically coherent moves, that represents a path of $\mathcal{T}$ that extends from $\mathcal{T}$'s root node $\mathcal{DM}_r$ to a leaf $\mathcal{DM}_l \in L_M$, in which each participant moves against its counterpart's last move.*

For example, in Figure 3.1 these are:

$$
\begin{aligned}
d_{(0\to2)} &=< \mathcal{DM}_0, \mathcal{DM}_1, \mathcal{DM}_2 > \\
d_{(0\to3)} &=< \mathcal{DM}_0, \mathcal{DM}_3 > \\
d_{(4\to5)} &=< \mathcal{DM}_4, \mathcal{DM}_5 > .
\end{aligned}
$$

However for convenience, in reference to a single dialogue tree we will refer to its disputes only with respect to their leaf indexes, i.e: $\mathcal{T} = \{d_1, \ldots, d_L\}$.

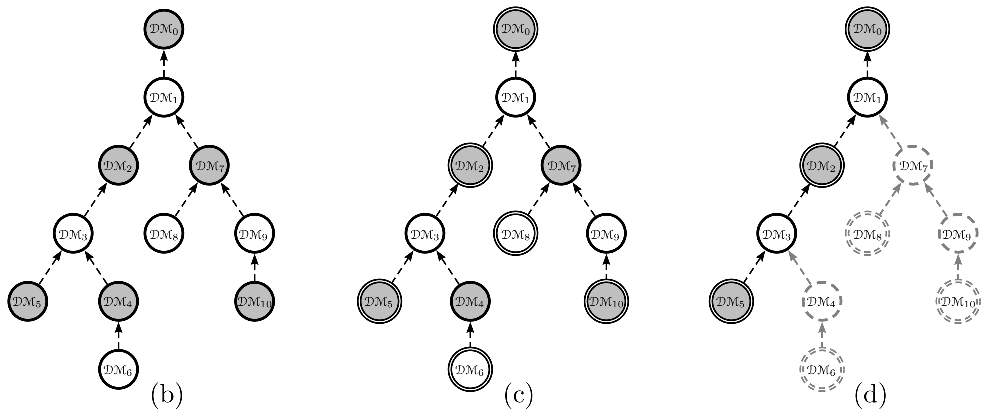
3.2.1.3 Defining the Winner

For defining the winner of the game we turn to Prakken [2005]'s work, who employs a labelling technique applied on a similar tree of locutions, instantiated from the dialogue process. The basic idea, is to assign a status to each of the tree's locutions, in order to eventually define the status of the claim in dispute. Intuitively, in order for Pr to fulfil its game role then for all the 'counter' moves moved

[1] Reasoning here is used in its informal form.

$\mathscr{D} =< \mathcal{DM}_0, \mathcal{DM}_1, \mathcal{DM}_2, \mathcal{DM}_3, \mathcal{DM}_4, \mathcal{DM}_5, \mathcal{DM}_6, \mathcal{DM}_7, \mathcal{DM}_8, \mathcal{DM}_9, \mathcal{DM}_{10} >$

(a)



(b) (c) (d)

Figure 3.2: a) A dialogue sequence, b) The corresponding dialogue tree $\mathcal{T}_0$ (Grey moves by Pr), c) A labelled dialogue tree, d) A *winning-strategy* for X, where $\mathcal{DM}_0^{con} = X$

into the game by Op, then Pr has to respond by attacking each one of them back. The latter captures the 'reinstatement principle' as it is expressed by Baroni and Giacomin [2009], while it also relates to the legal labellings proposed by Modgil and Caminada [2009] for their argument games.

Accordingly, in the context of a dialogue game this principle can be captured through the employment of the following definition:

Definition 33 (Labelling) *Assuming a legal persuasion dialogue $\mathscr{D}$ and a corresponding dialogue forest $\mathscr{F}$, then a node of a tree $\mathcal{T} \in \mathscr{F}$ is:*

- *labelled* ***in*** *iff all of its children-nodes are labelled* ***out***
- *labelled* ***out*** *iff it has at least one child labelled* ***in***

In the example depicted in Figure 3.2c we represent the `in` labelled nodes with a double-lined circle and the `out` labelled nodes with a single-dim-lined circle. This convention is followed throughout the thesis.

The proposed labelling-definition effectively reflects any game's semantics. In other words, since it is the case that a dialogue tree is implicitly instantiated through the dialogue process, it reflects the rules of the dialogue game. These rules encode restrictions on the moves that a participant can legally introduce in order to attack its interlocutor's moves, which in turn vary according to the semantics of interest. In this sense, one could provide an any time winning definition that would be based on the labelling status of the argument in dispute—the root argument in a dialogue tree—deciding on the winner according to whether $\mathcal{DM}_r$ is `in` or `out`, i.e. if $\mathcal{DM}_r$ is `in` then Pr is the winner, otherwise Op is.

Nevertheless, this definition cannot adequately reflect the *soundness and completeness properties* that should characterise the proposed dialogue system. As this will be further explained in Section 3.2.2, if a dialogue game is to be a proof of acceptance, then Pr has to attempt to build an admissible set of arguments. In other words, Pr's winning must be reflected by the existence of such a set amongst the branches of a dialogue tree. The latter cannot be captured by Prakken's labelling technique in the sense that even though one could decide on Pr currently being the winner of the game based on the labelling status of a root argument A, it is still possible that A may not be a member of an admissible extension in the implicitly instantiated AF.

We will, therefore, rely on this labelling technique only for relevance reasons, and resort to the notion of *winning-strategy*, as the latter is defined in Cayrol et al. [2003]; Modgil and Caminada [2009] but modified with respect to our framework, for providing an any-time winner definition.

Definition 34 (Winning-strategy) *Given a dialogue tree $\mathcal{T}$ then $\mathcal{T}'$ is a winning-strategy for X, iff:*

1. *$\mathcal{T}'$ is a finite sub-tree of $\mathcal{T}$, where X is the content of its root dialogue move;*

2. *Each dispute $d \in \mathcal{T}'$ ends with a move introduced by Pr;*

3. *$\forall d \in \mathcal{T}'$, given a sub-dispute d' of d whose last move is a move $\mathcal{DM}_{Pr}$ played by Pr, then for any $\mathcal{DM}_{Op} \in \mathcal{T}$ which extends a d' there exists a $d'' \in \mathcal{T}'$ such that d'–$\mathcal{DM}_{Op}$ is a sub-dispute of d'', and;*

4. *No two arguments X, Y moved by Pr in T' attack each other (i.e., the arguments moved by Pr are conflict-free).*

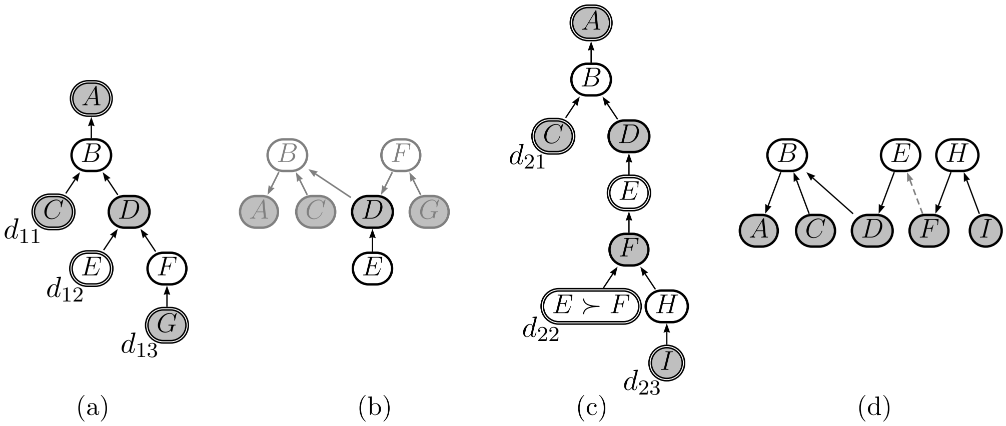


Figure 3.3: a) Tree $\mathcal{T}_1$ (Grey moves by Pr), b) The AF for $\mathcal{T}_1$, c) Tree $\mathcal{T}_2$, d) The AF for $\mathcal{T}_2$

The added value of this definition is that it imposes additional restrictions that can better reflect the acceptability status of Pr's arguments in the implicitly instantiated AF. That is whether they do or do not belong in a certain extension of AF, defined by the semantics of interest, according to whether Pr is currently winning respectively losing.

- Condition 1 defines the form of a winning strategy ($\mathcal{T}'$) as a finite sub-tree of a dialogue tree.

- Condition 2, reflects the reinstatement principle as it requires that all disputes that belong in a $\mathcal{T}'$ end with a Pr move. It is easy to see that it is otherwise possible for the labelling status of the root argument in $\mathcal{T}'$ to be labelled `out`.

- Condition 3, *partly* guarantees that all arguments that belong in $\mathcal{T}'$ moved by Pr form an admissible extension in the concurrently instantiated AF. So as to better understand the importance of this condition let us illustrate with an example.

In Figures 3.3a, 3.3b and 3.3c, 3.3d we respectively see two dialogue trees and their corresponding AFs. The disputes in the two trees are:

$$\begin{aligned}\mathcal{T}_1 &= \{d_{11}, d_{12}, d_{13}\} \\ &= \{< A, B, C >, < A, B, D, E >, < A, B, D, F, G >\} \\ \mathcal{T}_2 &= \{d_{21}, d_{22}, d_{23}\} \\ &= \{< A, B, C >, < A, B, D, E, F, (E > F) >< A, B, D, E, F, H, I >\}\end{aligned}$$

Note that the induced AFs are produced based on the defeat relations between the arguments and not based on their attack relations (we will further discuss how the defeat relation between the introduced arguments can be decided in the next section). If we only relied on the first two conditions then in the first case, illustrated in Figure 3.3a, we would decide that disputes d_{11} and d_{13} form a $\mathcal{T}'_1$ since they both end with a Pr move. Nevertheless, it is easy to see that in the corresponding AF (Figure 3.3b) argument D is not supported against E and thus cannot be part of what should be an admissible extension $\{A, C, D, G\}$. Similarly, in the second case depicted in Figure 3.3c, we would decide that since d_{21} and d_{23} end with a Pr move, they thus form a winning strategy. However, in the corresponding AF (Figure 3.3d) one can observe that F's attack on E does not succeed as defeat, as it is undermined by the preference move $E \succ F$, thus deeming argument D again not part of an admissible extension. In both cases, based on condition 3, since disputes d_{12} and d_{22} are not sub-disputes of some dispute $d \in \mathcal{T}'$, then d_{13} and d_{23} are discarded from $\mathcal{T}'$.

However, even with the addition of condition 3, the winning strategy definition alone cannot guarantee the soundness of the proposed framework with respect to any time termination, since, as it will be further discussed in the following section, it fails to account for an important aspect related with whether a particular move in a game *could have been repeated but has not*. The later refers to a considerable difference between the instantiated dialogue tree and the implicitly constructed AF, and is related with whether the winning status of the first can adequately reflect the acceptability properties of the other. We will however engage in further explaining this in the following sections.

Finally, the last condition reflects a basic property of any admissible set, which by definition is required to be conflict-free. We will henceforth refer to this property as the *conflict free-ness property.* In Figure 3.2d we can see the winning-strategy $\mathcal{T}'_0$ for $\mathcal{T}_0$ depicted in 3.2b. The existence of conflicting relationships between the moves employed by Pr is not apparent in the tree structure and is something that must be separately investigated. Moreover, it should be noted that the conflict free-ness property is limited to conflicts between the arguments employed by Pr since the preference-orderings of a participant are by definition consistent, and thus the appearance of a pre-orderings conflict between the moves of Pr is not possible.

Thus, at any point of a dialogue game we can decide on the winner of the game based on identifying a winning strategy in a $\mathcal{T} \in \mathscr{F}$.

Definition 35 *Let $\mathscr{D}$ be a persuasion dialogue and $\mathscr{F} = \{\mathcal{T}_0, \ldots, \mathcal{T}_{f-1}\}$ the induced dialogue forest, then at any point of the dialogue Pr is said to be the winner iff there exists at least one* winning-strategy *for a $\mathcal{T}_r \in \mathscr{F}$, for $1 \leq r \leq f-1$, else Op is the winner.*

The correlation between the labelling status of a dialogue tree and the existence of a winning strategy can be captured through the following propositions:

Proposition 1 *Let $\mathcal{T}$ be a dialogue tree and $\mathcal{DM}_0$ be $\mathcal{T}$'s root move, if there exists a $\mathcal{T}' \in \mathcal{T}$ then $\mathcal{DM}_0$ must be labelled* ***in****.*

Proof We prove this by contradiction. As an abuse of notation, let us write $< I, \mathcal{DM} >$ for a dialogue move introduced by $I \in \{Pr, Op\}$, while the extension of a dispute d_i with a dialogue move $\mathcal{DM}$ is denoted by the juxtaposition $d_i. < I, \mathcal{DM} >$.

Let there be a dialogue tree $\mathcal{T}$ and a winning strategy $\mathcal{T}' \in \mathcal{T}$, while the root move $< Pr, \mathcal{DM}_0 >$ is labelled `out`, then:

1. Based on Definition 33, if $< Pr, \mathcal{DM}_0 >$ is `out` then at least one of its attackers $(Op, \mathcal{DM}_j)$ must be labelled `in`.

2. As $< Pr, \mathcal{DM}_0 >$ is $\mathcal{T}$'s root move, we can assume that $d_0 = \mathcal{DM}_0$ is a sub-dispute of some dispute $d \in \mathcal{T}'$.

3. According to condition 3 of Definition 34, $d_0. < Op, \mathcal{DM}_j >$ must then be a sub-dispute of some $d'' \in \mathcal{T}'$.

4. Let us then assume a set of moves $\{\mathcal{DM}_{i_1}, \mathcal{DM}_{i_2}, \ldots, \mathcal{DM}_{i_k}\}$ which are moved into the game by Pr against $< Op, \mathcal{DM}_j >$.

5. Since $< Op, \mathcal{DM}_j >$ is `in`, it must be that all of its attackers are labelled `out`.

6. Accordingly, every such Pr move should at least have one attacker—an Op's move, e.g. $\mathcal{DM}_{j'}$—labelled `in`.

7. As we assume $d_0. < Op, \mathcal{DM}_j >$ to be a sub-dispute of some $d'' \in \mathcal{T}'$, inductively this must also be the case for when $d_0. < Op, \mathcal{DM}_j >$ is extended to:
$$d_0. < Op, \mathcal{DM}_j > . < Pr, \mathcal{DM}_{ik} > . < Op, \mathcal{DM}_{j'} >$$

8. Assuming that one extends the current dispute through repeating steps 3,4,5,6 and 7, substituting $\mathcal{DM}_j$ for $\mathcal{DM}_{j'}$ the result will simply be the construction of a longer dispute whose last argument will still be an Op's move labelled `in`.

9. Evidently, as we assume $\mathcal{T}$ to be finite, the construction of a dispute that ends with a Pr's move which will thus be part of $\mathcal{T}$'s winning strategy $\mathcal{T}'$ is impossible, which contradicts with condition 2 of the winning strategy definition (Definition 34).

Note at the same time, that it is not necessarily the case that when a root dialogue move is labelled `in` then there exists a winning-strategy in $\mathcal{T}$. This is formally expressed in the following proposition.

Proposition 2 *Let $\mathcal{T}$ be a dialogue tree and $\mathcal{DM}_0$ be $\mathcal{T}$'s root move, if $\mathcal{DM}_0$ is labelled* ***in*** *then it is not necessary that $\exists \mathcal{T}' \in \mathcal{T}$.*

The proof of this proposition relies on the use of a counter-example borrowed from the work of Modgil and Caminada [2009] used for a similar issue in their work. The example is presented in the Appendix A.1. It however requires that the reader is first aware of the protocol rules related to the credulous game (Section 3.2.2.1).

3.2.1.4 A More Strict Set of Protocol Rules

A drawback of the core protocol rules ($\mathscr{P}_{CP}$) is that they allow for redundant moves to be introduced into the game, i.e. they allow the introduction of moves that do not affect the winning status of the game. This obviously affects the game's efficiency. Additionally, since both participants introduce dialogue moves which belong to their distinct knowledge base and are not part of a pre-existing AF (as is the case with argument games Modgil and Caminada [2009]) but implicitly and incrementally form a 'shared' AF, we cannot know whether either Pr or Op will fully respond to the dialogical objectives of their roles. In other words, it is impossible to know whether either Pr or Op have actually introduced into the game all moves that they could in order to counter their interlocutor.

Though it is not possible to impose restrictions that will result in the participants complying fully to their dialogical objectives (a player may prefer to flee the dispute rather than be truthful), one may employ a more rigid approach so as to increase a game's efficiency. This can be achieved through dictating that in every turn both Pr and Op introduce dialogue moves that can alter the winning status of the game. This can be achieved through the employment of a more strict set of rules, that concurrently prevent the introduction of 'surplus' moves, in the sense that such moves are unable to change the winning status of the game (Prakken [2006]). However, the latter does not imply that such moves should generally be forbidden from being introduced in a dialogue, since a participant's motive may simply be to stall its counterpart.

It is enough that we reform the core protocol rules 3.1.a) and 3.2.a), as follows:

Strict Rules:

- if $\texttt{Speaker}(\mathcal{DM}_i) = Pr$ then:

  3.1.a)iv) $\mathcal{DM}_i$ is a reply to some $\mathcal{DM}_j$, $j < i$, subject to 3.1.a)i,ii,iii), that results in changing the winning status of the game.

- if $\texttt{Speaker}(\mathcal{DM}_i) = Op$ then:

  3.2.a)iv) $\mathcal{DM}_i$ is a reply to some $\mathcal{DM}_j$, $j < i$, subject to 3.2.a)i,ii,iii), that results in changing the the winning status of the game.

The incorporation of these rules in the core protocol rules, will be denoted as $\mathscr{P}_S$. Note that it is not necessary to impose this restriction in case 3.1.b) as that case the refers to the instantiation of a new dialogue tree which by default changes the winning status of the game.

3.2.2 Credulous & Grounded Semantics

In order to provide the rules for the credulous and the grounded persuasion dialogue games, we draw from the work of Modgil and Caminada [2009]. We must stress once more, that an important difference between argument and dialogue games is that argument games rely on *predefined* argumentation frameworks. In a sense, the outcome of an argument game simply reflects the image of a pre-existing AF instantiated from that knowledge-base, whereas dialogue games define the form of an implicitly and incrementally constructed AF from information accumulated from the exchanged locutions during a dialogue. Thus in other words, while the acceptability of an argument in a monologue game is predefined by a corresponding, pre-existing AF, in a dialogue game it depends on the final form of the incrementally constructed AF. That is to say that an argument game is a reasoning procedure that simply and strictly reflects the structure of an AF.

Nevertheless, for the purpose of our work, we can still employ the same set of protocol rules for both the credulous and the grounded semantics, regardless of this difference. This is because all we are concerned with, is whether these rules can adequately reflect the objective of a persuasion game. Essentially, if one disregards this essential difference between argument and dialogue games, in both games the participants assume exactly the same roles, those of Pr and Op. In this sense the protocol rules of argument games conform to the objective of a persuasion dialogue game.

3.2.2.1 Legal Persuasion dialogues

In relation to the credulous semantics, which reflect a rather lenient level of acceptability, Pr is characterised by a more flexible set of conditions compared to Op. As in argument games (Modgil and Caminada [2009]), the basic differences between the credulous and the grounded dialogue games concern Pr's and Op's ability to repeat any of the previous moves that have already been moved in a $\mathcal{T}$. The grounded semantics places additional difficulty to Pr's efforts of justification, since it essentially demands the introduction of indisputable information (un-attacked argument). The latter is reflected by corresponding protocol rules, according to which Pr is forbidden from repeating any of its previous dialogue moves in contrast to Op. At the same time Op is allowed to repeat both its own as well as Pr's moves. Of course, this is not the case with the credulous semantics where a symmetric attack may be both employed and repeated by Pr, in order to justify the credulous acceptance of its thesis.

At this point it is necessary that we introduce some additional notation, with respect to the distinct sets of moves that may be introduced in a dialogue game by each of the participants:

- $I(\mathscr{D})$: the dialogue moves introduced into the game by $I \in \{Pr, Op\}$
- $I(\mathcal{T})$: the dialogue moves found in $\mathcal{T}$ introduced by $I \in \{Pr, Op\}$
- $I(\mathcal{T}')$: the dialogue moves found in $\mathcal{T}'$ introduced by $I \in \{Pr, Op\}$
- For a set of moves S:
  - $\mathcal{R}^+(S)$: a set of dialogue moves that attack S
  - $\mathcal{R}^-(S)$: a set of dialogue moves that are attacked by S
- $\mathcal{R}^{\circlearrowleft}$: a set of self-attacking moves

Given these, as far as efficiency issues are concerned[1], it makes sense to restrict Pr from introducing self-attacking moves as this would immediately violate the properties of the admissible set which Pr is trying to construct in the game. As [Dunne and Bench-Capon, 2003, Defn. 6, Thm. 4] explain, if a credulous game is to be a proof of acceptance, then Pr essentially attempts an admissible set of arguments. We also note that in their work Dunne and Bench-Capon [2003] also establish tight bounds on the number of moves Pr has to advance: the size

[1]This is a matter of efficiency as a redundant move in the case where Pr achieves to construct a winning strategy will simply not be included in $Pr(\mathcal{T}')$.

of the smallest (number of arguments) admissible set containing the disputed argument.

Additionally, in relation to the grounded game, one could also restrict Pr from introducing a move the content of which shares a symmetric attack relationship with the contents of any other dialogue move that is already part of $\mathscr{D}$. This is because allowing Pr to do so makes no difference, since, as it will latter become evident, under the grounded game's rules Op can always reply through employing the same dialogue move which we allow Pr to attack, in order to counter Pr.

Similar restrictions were originally employed in the work of Amgoud and Cayrol [2002a]; Dunne and Bench-Capon [2003]; Modgil and Caminada [2009]; Prakken and Sartor [1997] which in our case they could accordingly be reflected by imposing:

$$\mathcal{DM}_i \in \mathcal{R}^+(Pr(d_{r\to l})) \cup \mathcal{R}^-(Pr(d_{r\to l})) \cup \mathcal{R}^{\circlearrowleft} = POSS(d_{r\to l})$$

Nevertheless, as it is further explained in Modgil and Caminada [2009], in the credulous game, even with imposing this restriction, it is still necessary to additionally check whether the moves introduced by Pr in a winning-strategy are conflict-free. This is because the restriction only concerns Pr's moves in each distinct line of dispute of a dialogue tree, and not Pr's moves in the whole tree, i.e. consistency across branches is not guaranteed.

While the latter is not possible in the grounded game[1], in the credulous game it is possible that a dialogue tree is produced $\mathcal{T}$ for which there might *seemingly* exist a winning-strategy $\mathcal{T}'$ for Pr, while though $Pr(\mathcal{T}')$ is not conflict-free[2]. This is suggested by proposition 2 introduced earlier with respect to the winning strategy.

This is therefore the reason why the winning definition provided in Section 3.2.1.3 is additionally characterised by the last condition, i.e. so as to avoid the appearance of a possible Pr victory that would falsely result in deciding that the argument in dispute is part of an admissible extension of the framework instantiated from the dialogue process. In other words, as is the case with the

[1]Refer to Modgil and Caminada [2009] for a justification of this claim.
[2]Refer to [Modgil and Caminada, 2009, p. 20] for an example.

grounded game, imposing the $POSS(d_{r\to l})$ protocol restriction in the credulous set of rules, will simply increase the game's efficiency but it will not guarantee the creation of a conflict-free set on behalf of Pr.

We also note that one could also restrict Op from introducing moves that can be attacked by any move found in $Pr(\mathscr{D})$, since, according to the credulous game rules, such a move cannot essentially change the final outcome of the game. For the purpose of our work, we disregard efficiency issues, and focus only on applying two sets of basic restrictions for differentiating between the credulous and the grounded semantics. We do so by augmenting the strict rules $\mathscr{P}_S$ with rules that place restrictions on Op's, respectively Pr's moves, as follows:

Credulous Rules: Let $\mathcal{T}_{\mathscr{D}}$ be a dialogue induced for a dialogue $\mathscr{D}$ then:

- if $\texttt{Speaker}(\mathcal{DM}_i) = Pr$ then:

  3.1.a)v) $\forall d \in \mathcal{T}_{\mathscr{D}}$, $\nexists \mathcal{DM}_j \in d$ where:
  $\texttt{Speaker}(\mathcal{DM}_j) = Pr$ and $\mathcal{DM}_i^{con}$=$\mathcal{DM}_j^{con}$
  (i.e., Pr cannot repeat locutions in any given dispute).

Grounded Rules: Let $\mathcal{T}_{\mathscr{D}}$ be a dialogue induced for a dialogue $\mathscr{D}$ then:

- if $\texttt{Speaker}(\mathcal{DM}_i) = Op$ then:

  3.2.a)v) $\forall d \in \mathcal{T}_{\mathscr{D}}$, $\nexists \mathcal{DM}_j \in d$ where:
  $\texttt{Speaker}(\mathcal{DM}_j) = Op$ and $\mathcal{DM}_i^{con}$=$\mathcal{DM}_j^{con}$
  (i.e., Op cannot repeat locutions in any given dispute).

Henceforth, $\mathscr{P}_{GS}$ will denote the protocol defined by the credulous rules, and $\mathscr{P}_{CS}$ the protocol defined by the grounded rules. Furthermore, we will refer to dialogues that comply with the credulous or the grounded protocol rules as *legal* credulous or grounded persuasion dialogues respectively.

A trivial example on both types of games that relies on a symmetric attack relationship between two dialogue moves is shown in Figure 3.4. For convenience, we don't alter the move indexes in order to stress repetition. It is easy to notice that in the credulous games depicted in Figure 3.4a, given the symmetric

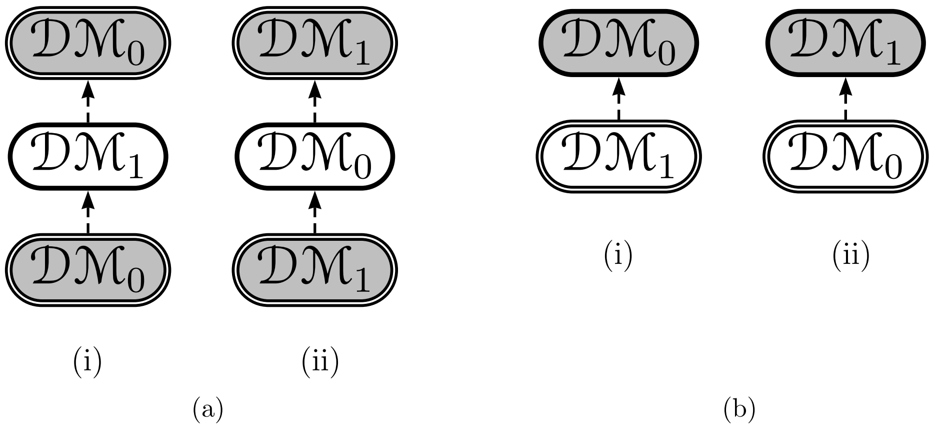


Figure 3.4: a) The credulous games (Pr is grey), b) and the grounded games for the moves $\mathcal{DM}_0$ and $\mathcal{DM}_1$

attack relationship between the dialogue moves $\mathcal{DM}_0$ and $\mathcal{DM}_1$, the conclusions of the containing arguments of these moves that serve as the claims in dispute respectively, are both deemed acceptable under the credulous semantics. On the contrary in the grounded dialogue games (Figure 3.4b) none of these claims are be justified.

Based on these sets of protocol restrictions with respect to a semantics:

$$\mathcal{S} = \{\mathscr{P}_{CS}, \mathscr{P}_{GS}\}$$

respectively representing the Credulous and the Grounded semantics, we are able to prove the following lemma:

Lemma 1 *At any state of a finite legal $\mathcal{S}$-persuasion dialogue $\mathscr{D}$, there can only exist at most one dialogue tree $\mathcal{T}_{\mathcal{DM}_k}$, where the dialectical status of its root move is labelled* `in`*.*

Proof Let us represent the state of a dialogue game in a turn k as two sets, $(IN,\ OUT)_k$, where IN contains the trees with a root move labelled `in`, and OUT contains the trees with a root move labelled `out`. Additionally, let the total number of trees instantiated in a dialogue game at a given state be expressed as

$|\mathscr{F}^{\mathscr{D}}| = N$, for which it also holds that $N = |IN| + |OUT|$, while it is also evident that at any point of the game if $|IN| = n$, then $|OUT| = N - n$. Moreover, let $(|IN|, |OUT|)_k$ represent the number of trees in each set in a given turn.

Following on, we remind the reader that the $\mathscr{P}_S$ protocol, that characterises both the $\mathscr{P}_{GS}$ and the $\mathscr{P}_{CS}$ protocols, dictates that either of the participants can introduce a new dialogue move only if it changes the dialectical status of a root move. Moreover, in the course of a dialogue game, based on rule 3.1.b), only Pr is capable of instantiating a new dialogue tree through introducing a new dialogue move, comprising a different argument for c_{init}, which can be perceived as the trivial case where the labelling status of a root move is 'changed' to `in`. In this sense legal transitions between the different states of a dialogue game can be expressed as follows, with respect to whether it is Pr's or Op's turn to move:

1. $(n, N-n) \xrightarrow{Pr} (n+1, N-n-1)$

- Pr introduces a move into the current dialogue tree against Op's last move
- Pr introduces a move through backtracking to a previous point into the current dialogue tree
- Pr introduces a move through backtracking to a previous point into a different dialogue tree
- Pr introduces a new root move, and instantiates a new dialogue tree

2. $(n, N-n) \xrightarrow{Op} (n-1, N-n+1)$

- Op introduces a move into the current dialogue tree against Pr's last move
- Op introduces a move through backtracking to a previous point into the current dialogue tree
- Op introduces a move through backtracking to a previous point into a different dialogue tree

Given these, it can be trivially shown that since Pr makes the first move, for turn $k = 0$, the state of the dialogue game is $(\{\mathcal{T}_{\mathbb{DM}_0}\}, \{\varnothing\})_{k=0}$, thus $(1, 0)_{k=0}$. Let us now assume that in a turn k there exist $|IN| = n > 1$ dialogue trees. We then differentiate between the following two cases:

- *Pr moves in k*: This implies that in turn $k-1$, $|IN| = n-1$, based on legal transition 2 since it was *Op*'s turn to move. In turn, the latter implies that in $k-2$, $|IN| = (n-1)+1 = n$, based on legal transition 1. Progressing accordingly, given that the dialogue is finite, and given that, based on the protocol rules, P makes the first move, then for $k = 0$ it is definitely the case that $|IN| = n$, thus $(n, 0)_{k=0}$. Nevertheless, as we have already shown for $k = 0$, it holds that $(1, 0)_{k=0}$, which contradicts with the assumption that $n > 1$.

- *Op moves in k*: This implies that in turn $k-1$, $|IN| = n+1$, based on legal transition 1 since it was *Pr*'s turn to move. In turn, the latter implies that in $k-2$, $|IN| = (n+1)-1 = n$, based on legal transition 2. Progressing accordingly, given that the dialogue is finite, and given that, based on the protocol rules, P makes the first move, then for $k = 0$ it is definitely the case that $|IN| = n+1$, thus $(n+1, 0)_{k=0}$. Nevertheless, as we have already shown for $k = 0$, it holds that $(1, 0)_{k=0}$, which contradicts with the assumption that $n > 1$.

Generally, we have shown that given a protocol characterised by the $\mathscr{P}_S$ protocol, at any state of the dialogue game it is impossible for $|IN| > 1$.

3.3 Soundness & Fairness

In this Section we focus on evaluating the proposed dialogue system for persuasion dialogues with respect to its *soundness* and its *fairness*. As explained by Prakken [2005], the notion of soundness serves as a characterisation that concerns the level of correspondence between the any-time winning definition of a dialogue, and the underlying logic that concerns the accumulated information gathered from the dialogue process. In our case, based on the any-time winning definition provided in Section 3.2.1.3, at any point of the dialogue we can decide that either *Pr* or *Op* wins, based on the existence of a winning-strategy. Then, in order for the outcome of the game to be sound one has to prove that:

- in the case where *Pr* currently wins then the claim in dispute is *justified* under the game's semantics and with respect to the logical content accumulated from the dialogue process, or;

- in the case where *Op* currently wins then the claim in dispute is *not justified* under the game's semantics and with respect to the logical content accumulated from the dialogue process.

In other words, proving the soundness of the proposed system relies on proving that the any-time winning definition, which essentially decides the current acceptability status of the claim in dispute, can be reflected by the construction of a *justified* argument from the logical content obtained from the dialogue moves that were introduced into the dialogue game. The notion of fairness concerns proving the same thing, but in reverse. That is to prove that assuming that a justified argument for the claim in dispute can be constructed from the currently agreed upon information, then there exists a winning strategy for that argument.

3.3.1 Conflicts between Preference-orderings

The basic idea is to prove a relation between the any-time winning definition and the underlying logic. To recap, both *Pr* and *Op* can move arguments and preferences constructed from their own argumentation theories and the commitment stores of their interlocutor. The latter contains all the content exchanged through locutions. Our aim is to prove soundness and fairness with respect to the Dung framework $AF_{\mathscr{D}} = (\mathcal{A}^{\mathscr{D}}, \mathcal{D})$ of arguments ($\mathcal{A}^{\mathscr{D}}$) and defeats ($\mathcal{D}$) that is defined by the accumulated knowledge in the commitment stores. We assume this $AF_{\mathscr{D}}$ to be implicitly instantiated from the dialogue process. The defeat relation between the introduced arguments can be defined with respect to the type of attacks that characterise the attack relationships between the employed arguments, and the preference-orderings moved into the game according to Definition 15.

However, we first need to account for the possibly conflicting preference information introduced into the game by the participants. In other words, to decide in cases were conflicting preferences information is introduced by the two participants, which should be discarded.

Let us suppose that the following preferences on two arguments X and Y, i.e. the preference-orderings on rules and premises defining these preferences, are moved into a game (and thus included in each of the participants' commitment stores) as follows: Pr moves $X \prec Y$ followed by a counter preference-ordering $Y \prec X$ moved by Op. Relying on the rules of the grounded game, it can be easily seen that Op's preference will win out over Pr's preference, while respectively in the case of the credulous game, Pr's preference will win out over Op's preference. Intuitively, consider these conflicting preferences as mutually attacking arguments with no other incoming attacks. Then Pr's argument (preference) will be justified under the credulous, but not grounded semantics.

We begin by providing the auxiliary components of the implicitly instantiated $AF_{\mathscr{D}}$ framework, expressed as components of a preference-based argumentation framework (PAF). For doing so we first provide a more detailed version of definition 23 which encompasses four distinct update mechanisms with respect to each of the logical components $\mathcal{K}, \leq', \mathcal{R}, \leq$ found in a sub-theory S of a participant's AgT.

Definition 36 (*ASPIC*$^+$ Commitment Stores) *Given a set of agents* $\{Ag_1, \ldots, Ag_\nu\}$ *participating in a dialogue* $\mathscr{D} = < \mathcal{DM}_0, \ldots, \mathcal{DM}_n >$*, for any agent* Ag_i *a* commitment store *is a tuple* $CS_i =< \mathcal{K}i, \leq' i, \mathcal{R}i, \leq i >$ *whose evolution we define as follows:*

1. $CSi^0 = < \varnothing, \varnothing, \varnothing, \varnothing >$

2. For $k = 0 \ldots n$*,* CSi^{k+1} *is obtained by updating* CSi^k *with the contents of the dialogue move* $\mathcal{DM}_k^{con} = X$ *based on the following definition of an update function* ***update****:*

$$\texttt{update}(CSi^k, \mathcal{DM}_k) \longrightarrow \begin{cases} CSi^{k+1} =< \mathcal{K}i^k \cup \texttt{Prem}(X), \leq' i^k, \mathcal{R}i^k \cup \texttt{Rules}(X), \leq i^k > & (a) \\ CSi^{k+1} =< \mathcal{K}i^k, \leq' i^k \cup X, \mathcal{R}i^k, \leq i^k > & (b) \\ CSi^{k+1} =< \mathcal{K}i^k, \leq' i^k, \mathcal{R}i^k, \leq i^k \cup X > & (c) \end{cases}$$

$$\begin{array}{ll} \textit{if } X \in \mathcal{A} & (a) \\ \textit{if } X \in \leq' i & (b) \\ \textit{if } X \in \leq i & (c) \end{array}$$

Note that we will henceforth refer to the credulous and the grounded semantics through $\mathcal{S} = \{\text{Credulous, Grounded}\}$. Based on the contents of the commitment stores of the participants, we define the components of the implicitly instantiated $AF_{\mathscr{D}}$ as follows:

Definition 37 **(PAF)** *Assume a legal $\mathcal{S}$-persuasion dialogue $\mathscr{D}$, where CS_{Pr} and CS_{Op} are respectively Pr's and Op's commitment stores. Let S be a tuple of the form $< \mathcal{K}, \leq', \mathcal{R}, \leq >$, and* $\texttt{Args(S)}$*,* $\texttt{Prefs(S)}$*, and* $\texttt{Attacks(S)}$ *be three functions that respectively return: the arguments that may be constructed from S; the preference-orderings on those arguments, and; the attacks between those arguments, then one may express its logical contents as a triple $< \mathcal{A}^{\mathscr{D}}, \mathcal{C}^{\mathscr{D}}, \mathcal{P}^{\mathscr{D}} >$ where:*

- $\mathcal{A}^{\mathscr{D}} = \texttt{Args}(CS_{Pr} \cup CS_{Op})$
- $\mathcal{C}^{\mathscr{D}} = \texttt{Attacks}(\mathcal{A}^{\mathscr{D}})$
- $\mathcal{P}^{\mathscr{D}} = \mathcal{P}^{Pr} \circledast \mathcal{P}^{Op}$*, where* $\mathcal{P}^{Pr} = \texttt{Prefs}(CS_{Pr})$ *and* $\mathcal{P}^{Op} = \texttt{Prefs}(CS_{Op})$ *such that:*
  - $\mathcal{P}^{Pr} \circledast \mathcal{P}^{Op} = \mathcal{P}^{Pr} \cup \{(X > Y) | (X > Y) \in \mathcal{P}^{Op}, (Y > X) \notin \mathcal{P}^{Pr}\}$ *if* $\mathcal{S} = \mathscr{P}_{CS}$ *(Credulous)*
  - $\mathcal{P}^{Pr} \circledast \mathcal{P}^{Op} = \mathcal{P}^{Op} \cup \{(X > Y) | (X > Y) \in \mathcal{P}^{Pr}, (Y > X) \notin \mathcal{P}^{Op}\}$ *if* $\mathcal{S} = \mathscr{P}_{GS}$ *(Grounded)*

  where $\mathcal{P}^{Pr}$ and $\mathcal{P}^{Op}$ represent the preference-orderings moved into the game by Pr respectively Op, while we assume transitive closure for $\mathcal{P}^{\mathscr{D}}$.

In relation to the `Attacks` function, we note that it relies on a contrariness function $^{-}$ applied on $\mathcal{L}$ for inducing the binary attack relations between the elements of $\mathcal{A}^{\mathscr{D}}$. However, for convenience as an abuse of notation we will assume that `Attacks` is simply applied on $\mathcal{A}^{\mathscr{D}}$. Furthermore in relation to set $\mathcal{A}^{\mathscr{D}}$, we note that it does not necessarily represent the set of all arguments introduced into a dialogue game. To differentiate between these two sets we will hence refer to the set of arguments employed in a dialogue game as $\mathcal{A}'$, where $\mathcal{A}' \subseteq \mathcal{A}^{\mathscr{D}}$. This is due to the fact that additional arguments, not employed in the game, could be instantiated from the logical contents accumulated in both CS_{Pr} and CS_{Op}.

Similarly, we differentiate between the attack relationships in $\mathcal{A}^{\mathscr{D}}$ and $\mathcal{A}'$ with $\mathcal{C}^{\mathscr{D}}$ and $\mathcal{C}'$ respectively where $\mathcal{C}' \subseteq \mathcal{C}^{\mathscr{D}}$, as well as between the preference-orderings employed in the dialogue game and the actual set of preference-orderings with $\mathcal{P}'$ and $\mathcal{P}^{\mathscr{D}}$.

As far as $\mathcal{P}^{\mathscr{D}}$ is concerned, the possible conflicts between the distinct preference sets of the two participants, $\mathcal{P}^{Pr}$ and $\mathcal{P}^{Op}$ are simply resolved according to the game's semantics. Essentially, in the case where $\mathcal{S}$ =Credulous then:

- $\mathcal{P}^{\mathscr{D}}$ consists of the preference-orderings moved by Pr, maximally consistently extended with the preference-orderings moved by Op;

whereas in the case of $\mathcal{S}$ =Grounded:

- $\mathcal{P}^{\mathscr{D}}$ consists of the preference-orderings moved by Op, maximally consistently extended with the those moved by Pr.

In essence, what is actually discarded is the conflicting preference-orderings over non-axiom premises and defeasible rules, and *not* preference-orderings over arguments. However for convenience, we express the conflict resolution definition with respect to the preference-orderings over the arguments which result from the preference-orderings over the logical components that compose them.

3.3.2 Conditional Soundness

Having resolved the conflicts between preference-orderings introduced in a $\mathscr{D}$, we can instantiate an $AF_{\mathscr{D}}$ with respect to the decided defeat relation over $\mathcal{A}^{\mathscr{D}}$. That is to induce an $AF_{\mathscr{D}}$ from a PAF. At this point it is worth reminding the reader that, based on Definition 15, in the special case where there exists an unresolved preference-dependent attack relationship between two arguments X and Y, $(X, Y) \in \mathcal{C}^{\mathscr{D}}$, for which no preference information is introduced into the game, then it is implicitly suggested that $X \nprec Y'$ (as well as $Y \nprec X'$), which in turn implies that $(X, Y) \in \mathcal{D}$. As explained in Section 2.3.1, the extensions of $AF_{\mathscr{D}}$ can then be defined in a similar way to the extensions of a Dung framework.

In order to prove the soundness of the proposed framework, all we need to do is to compare all the possible winning outcomes of all the intermediate states of a dialogue process—including the initiating and the final state—with the corresponding $AF_{\mathscr{D}}$ frameworks instantiated for each of those states, and prove that given a semantics $\mathcal{S}$, whenever $Pr(Op)$ is the winner, then correspondingly there exists(does not exist) an argument contained in an `in` labelled root dialogue move $\mathcal{DM}_r$, which is(is not) part of the $\mathcal{S}$-extension in $AF_{\mathscr{D}}$.

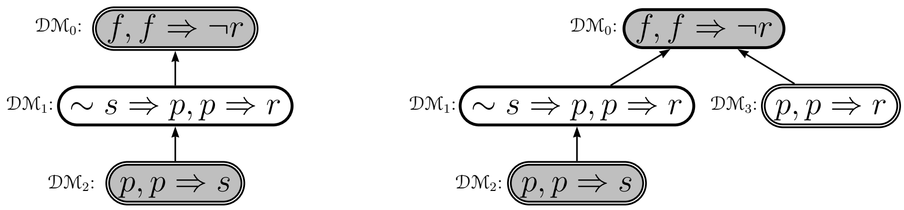
As trivial as this may seem, basic differences between a dialogue and the concurrently instantiated $AF_{\mathscr{D}}$ make this process complicated, while if one accounts for the underlying logic then more things need to be considered in order to guarantee the soundness of a dialogue's outcome. This is because, as it has already been stressed, a dialogue system which relies on an abstract analysis cannot take into account the possibility of a new argument being instantiated from the logical content accumulated from the dialogue process which could possibly alter the winning status of the game if introduced. In this sense, and as it is also discussed in Prakken [2001, 2005], a currently terminated dialogue might not yet be 'logically' terminated. Simply put, this possibility eludes an abstract analysis.

Essentially, the distinction between a "currently terminated" and "logically terminated" dialogue is one feature differentiating "argumentation" from "mathematical proof". As [Bench-Capon and Dunne, 2007, point (c), p. 620] explain, let us assume P to be a formal proof that T holds, then in mathematical reasoning:

> *Conclusions are final and definite: if P is a correct proof that T, then T is,* ipso facto *valid and this status does not admit subsequent qualification or amendment, let alone retraction.*

However, as they further explain, in argument and discussion as encountered in everyday contexts, it is rare this feature applies—assuming that one perceives for example P as a persuasive argument for accepting T.

Consider the example depicted in Figure 3.5 which concerns a grounded dialogue game between two participants. In Figure 3.5a, there obviously exists a winning strategy for $f, f \Rightarrow \neg r$ thus making Pr is the current winner. Particularly, the game concerns the acceptability of $\neg r$—the conclusion of the argument comprised by the tree's root move $\mathcal{DM}_0$. $\mathcal{DM}_1$ constitutes a rebut attack on $\mathcal{DM}_0$ since it infers r, which is a contradicting claim to $\neg r$. Pr then responds with another attack on the contents of $\mathcal{DM}_1$ with $\mathcal{DM}_2$, since the conclusion of its



(a) (b)

Figure 3.5: a) A logically-incomplete dialogue, b) A logically-complete dialogue

comprised argument contradicts with the premise of the sub-argument employed as a premise in $\mathcal{DM}_1$, in order to conclude r.

Nevertheless, as it is shown in Figure 3.5b, with the introduction of $\mathcal{DM}_2$ we can now assume that the p premise, inferred as a conclusion of the sub-argument $\sim s \Rightarrow p$ in $\mathcal{DM}_1$, no longer needs to be supported by $\sim s \Rightarrow p$ since it is now part of Pr's commitments after using it to infer s in $\mathcal{DM}_2$, i.e. by uttering $p, p \Rightarrow s$, Pr commits to believing p, and making Op aware p. Thus a new argument can now be instantiated on behalf of Op through the combination of p, and $p \Rightarrow r$ already introduced by Op, which can then be moved into the game as the content of a new move $\mathcal{DM}_3$, altering the dialectical status of the claim in dispute (Figure 3.5b).

Notice that in contrast with $\mathcal{DM}_1$, $\mathcal{DM}_3$ is not susceptible to an assumption attack (an attack on $\sim s$), since now the necessary premise for inferring r is directly provided through Pr's commitment store.

3.3.2.1 Logical & Protocol Completeness

As explained in Prakken [2005], accounting for logical completeness means that additional conditions must be satisfied with respect to whether at a current state of a dialogue game the *players have moved all arguments that they were allowed to move*, according to the game's protocol. Particularly, this restriction refers to two categories of arguments. The first concerns arguments which may be instantiated from the logical contents found in the commitment stores of both participants (as illustrated in Figure 3.5b), while the second concerns arguments which are already part of the dialogue (arguments that were previously introduced) and could possibly be repeated.

The necessity of imposing restrictions concerned with the first category of arguments is self-evident, as it is essential that the participants account for the possibility that new arguments could be instantiated and deployed in a game. However, additionally imposing conditions concerned with the second category of arguments has to be further analysed. In essence the latter results from the fact that in a dialogue game the instantiated dialogue forest and the concurrently instantiated $AF_{\mathscr{D}}$ are different in one considerable aspect. Given either the $\mathscr{P}_{CS}$ or the $\mathscr{P}_{GS}$ protocol, the winning status of the initiating move of a dialogue game changes with each consequent move introduced into the game (as imposed by the strict set of protocol rules $\mathscr{P}_S$). This is not reflected in the corresponding $< \mathcal{A}^{\mathscr{D}}, \mathcal{D} >$ framework with an analogous participation or non-participation of X in the $\mathcal{S}$-extension. Essentially, this is due to the fact that an argument can appear only once in an $(A^{\mathscr{D}}, \mathcal{D})$ framework. Thus while the repetition of a move is immediately reflected in a dialogue tree with the addition of a new move as a *leaf*—which, according to the game's rules, affects the winning status of the game—in contrast, the corresponding $AF_{\mathscr{D}}$ framework remains intact as do its extensions.

The latter implies that in some intermediate states of the dialogue game, the dialogue's winning status might not comply with a corresponding participation of X in the $\mathcal{S}$-extension. Thus, in general, guaranteeing an any time soundness for our framework is not possible based on the winning strategy definition alone, as the result for such states could be both *not sound* and *unfair*. Therefore, the soundness of the proposed system is additionally subjected to conditions related to the possible dialogue moves that have already been introduced and can be repeated in the game.

For example, in the trivial case of the credulous dialogue game depicted in Figure 3.6a, if the game is interrupted after the introduction of $\mathcal{DM}_0$—where in this case it is the second move introduced into the game—then we can decide that there does not exist a winning strategy for $\mathcal{DM}_1$ and accordingly decide that the argument comprised by $\mathcal{DM}_1$ is not acceptable under the credulous semantics. However, if we assume a symmetric attack relationship between the arguments comprised in moves $\mathcal{DM}_0^{con} = A$ and $\mathcal{DM}_1^{con} = B$ then under the credulous game protocol rules, Pr should be allowed to repeat $\mathcal{DM}_1$ while Op is not allowed to repeat $\mathcal{DM}_0$. At the same time, based on the current form of the game's

(a) (b)

Figure 3.6: a) A trivial case of a credulous game (Pr is grey), b) The corresponding AF.

dialogue tree there doesn't exist a winning strategy for the argument comprised by $\mathcal{DM}_1$, while it is easy to see that based on the form of the instantiated $AF_{\mathscr{D}}$ both arguments comprised by the dialogue moves in respect, should be acceptable under the credulous semantics (Figure 3.6b). In such a case, we say that the game is not *protocol complete*.

These two limitations of the winning strategy definition with respect to guaranteeing the soundness of the proposed framework, make necessary that we introduce two distinct sets of conditions concerned with the arguments and the preferences that may be repeated in a game.

Definition 38 (Protocol Completeness) *Given a legal $\mathcal{S}$-persuasion dialogue $\mathscr{D}$, let $\mathcal{T}$ be $\mathscr{D}$'s dialogue tree such that $d_1 \in \mathcal{T}$ and $d_2 \in \mathcal{T}$, we then say that $\mathscr{D}$ is protocol complete if:*

1. *$\exists X \in d_1$ and $Y \in d_2$, where $X, Y \in \mathcal{A}'$ then if Y can legally be moved against X in d_1 then it has been.*

2. *$\exists (X > Y) \in d_1$ and $(Y > X) \in d_2$, where $(X > Y), (Y > X) \in \mathcal{P}'$ then if $(Y > X)$ can legally be moved against $(X > Y)$ in d_1 then it has been.*

Definition 39 (Logical Completeness) *Given a legal $\mathcal{S}$-persuasion dialogue $\mathscr{D}$, we say that $\mathscr{D}$ is logically complete if:*

1. *$\exists X \in \mathcal{A}^{\mathscr{D}}$ then if X can legally be moved in $\mathscr{D}$ then it has been.*

2. *$\exists (X > Y) \in \mathcal{P}^{\mathscr{D}}$ then if $(X > Y)$ can legally be moved in $\mathscr{D}$ then it has been.*

Essentially, protocol completeness accounts for the possible legal repetitions of moves that have to be repeated into the game for it to be deemed sound, while logical completeness generally accounts for all possible arguments or preference-orderings that can be instantiated from the dialogue's accumulated logical content, that should also be moved into the game for the same reason. In this respect, logical completeness subsumes protocol completeness. Notice however that, in the case of preferences this does not hold. This is because the set of preference relations moved in an actual game are a superset of the preference relations defined by the commitment stores of the participants, as conflicting preference information is discarded according to Definition 37 and with respect to either $\mathscr{P}_{CS}$ or $\mathscr{P}_{GS}$. Nevertheless, it is still the case that if Pr wins/loses a logically complete dialogue then it is implied that Pr would also win/lose a protocol complete game. The only difference is simply that in the case of the logically complete game, we lose some information in relation to whether a preference-ordering in $\mathcal{P}^{\mathscr{D}}$ was actually countered with a preference-rebut attack in the protocol complete game, which could then be discarded. However, this has little meaning as the outcome in both cases with respect to the success of a certain attack as defeat is the same.

3.3.3 Soundness & Fairness Results

Having provided a definition for a logically-complete dialogue, we can now provide a number of theorems in relation to the soundness and fairness of any dialogue produced by the proposed system.

We first introduce some basic notation for the following proofs. We assume a persuasion dialogue $\mathscr{D}$, for a claim c_{init}, for which $\mathscr{F}$ and $AF_{\mathscr{D}} = (\mathcal{A}^{\mathscr{D}}, \mathcal{D})$ are the dialogue forest respectively the argumentation framework defined by $\mathscr{D}$. Let $AF_{\mathscr{D}}$ be induced by a $PAF =< \mathcal{A}^{\mathscr{D}}, \mathcal{P}^{\mathscr{D}}, \mathcal{C}^{\mathscr{D}} >$ as the latter is described in Definition 37. Let $T \in \mathscr{F}$ be a tree in $\mathscr{F}$, the content of the root move of which is an argument X for c_{init}, with a winning strategy $\mathcal{T}' \in \mathcal{T}$ for Pr, where each $d_i(d'_i)$ is of the form $< \mathcal{DM}_r, \ldots, \mathcal{DM}_\tau >$ for a root dialogue move $\mathcal{DM}_r$ and for a terminal move (leaf node) $\mathcal{DM}_\tau$, while we assume $\{d_1, \ldots, d_n\}$ and $\{d'_1, \ldots, d'_m\}$ are the disputes in $\mathcal{T}$ respectively $\mathcal{T}'$.

As an abuse of notation, we write $I - \mathcal{DM}^{con} / I - \mathcal{DM}$ for a dialogue move

$\mathcal{DM}$ introduced by $I \in \{Pr, Op\}$, i.e. $Pr - X/Pr - \mathcal{DM}$, while the extension of a dispute d with a dialogue move $\mathcal{DM}$ is denoted by the juxtaposition $d_i, I - \mathcal{DM}^{con}/d_i, I - \mathcal{DM}$. For example disputes will appear in the following forms:

$$d_{i'} : Pr - \mathcal{DM}_i, Op - \mathcal{DM}_j, Pr - \mathcal{DM}_{i'}$$

$$d_{i'} : Pr - X, Op - Y, Pr - Z$$

while extended disputes will appear as follows:

$$d_{k'} : d_{i'}, Pr - \mathcal{DM}_k, Op - \mathcal{DM}_l, Pr - \mathcal{DM}_{k'}$$

$$d_{k'} : d_{i'}, Pr - A, Op - B, Pr - C$$

In addition, if applicable, the dispute's index will be consistent with the index of its last move.

The set of all arguments moved in a tree $\mathcal{T}$ or in a winning strategy $\mathcal{T}'$ by I will be denoted as $I(\mathcal{T})$ respectively $I(\mathcal{T}')$. Furthermore, though a preference-ordering between two arguments X and Y is expressed as a pair $(X, Y) \in \mathcal{P}^{\mathscr{D}}$ which implies that X is preferred over Y, for convenience we will express this preference relationship as $X > Y$, though, strictly following the $ASPIC^+$ notation, $X > Y$ implies that X is strictly preferred over Y.

Theorem 1 (Credulous Soundness) *Let $AF_{\mathscr{D}} = (A^{\mathscr{D}}, \mathcal{D})$ be the argumentation framework defined by a finite legal logically-complete credulous persuasion dialogue $\mathscr{D}$ for a claim c_{init}. Then if Pr is winning, there exists an argument X for c_{init} such that X is in the admissible extension of $AF_{\mathscr{D}}$.*

Proof Assume a finite legal logically-complete credulous persuasion dialogue $\mathscr{D}$ for a claim c_{init} where Pr is currently winning. Then based on Definition 35 and Lemma 1 there must exist a $\mathcal{T} \in \mathscr{F}$, the content of the root move of which is an argument X for c_{init}, with a winning strategy $\mathcal{T}' \in \mathcal{T}$ for Pr.

Then X's membership in an admissible extension of $AF_{\mathscr{D}}$ can be shown if $Pr(\mathcal{T}')$ is admissible. That is to prove that:

1. $Pr(\mathcal{T}')$ is conflict free and;

2. $\forall A \in Pr(\mathcal{T}')$ if $\exists B \in \mathcal{A}^{\mathscr{D}}$ such that $(B, A) \in \mathcal{D}$ then $\exists C \in Pr(\mathcal{T}')$ such that $(C, B) \in \mathcal{D}$.

The property of conflict free-ness is accounted by means of condition 4 of Definition 34. We prove that the second property holds, by contradiction.

Assume that $Pr(T')$ is not admissible. This implies that for some $A \in Pr(\mathcal{T}')$ introduced by a move $< Pr, \mathcal{DM}_i >$ there must exist an argument $B \in \mathcal{A}^{\mathscr{D}}$ such that the following conditions hold:

1. $(B, A) \in \mathcal{C}^{\mathscr{D}}$
2. $(B, A) \in \mathcal{D}$
3. $\nexists C \in Pr(\mathcal{T}')$ such that $(C, B) \in \mathcal{D}$

Suppose $\mathcal{DM}_i \in d'_l$ of a $\mathcal{T}'$ such that $d'_l : Pr - \mathcal{DM}_r, \ldots, Pr - \mathcal{DM}_i$. Since $A \in Pr(\mathcal{T}')$ and $(B, A) \in \mathcal{C}^{\mathscr{D}}$, then by completeness conditions B must have been introduced into $\mathscr{D}$ as the content of an Op move $(\mathcal{DM}_j)$, appearing: either directly after $Pr - \mathcal{DM}_i$ in d'_l as illustrated in Figure 3.7, or; if this violates the repetition conditions imposed by the protocol rules for the credulous game $(\mathscr{P}_{CG})$ then it must be that $Op - \mathcal{DM}_j$ appears before $Pr - \mathcal{DM}_i$ in d'_l as illustrated in Figure 3.8 (we remind the reader that, based on $\mathscr{P}_{CG}$, Op cannot repeat a move in the same dispute, while Pr can). Suppose that I) is the case. Then:

$$d'_l : \boxed{\mathcal{DM}_r} \longleftarrow \ldots \longleftarrow \boxed{\mathcal{DM}_i^{con} = A} \longleftarrow \boxed{\mathcal{DM}_j^{con} = B}$$

Figure 3.7: Case I

$$d'_l : \boxed{\mathcal{DM}_r} \longleftarrow \ldots \longleftarrow \boxed{\mathcal{DM}_j^{con} = B} \longleftarrow \ldots \longleftarrow \boxed{\mathcal{DM}_i^{con} = A}$$

Figure 3.8: Case II

a) By definition of the winning strategy, d'_l cannot end with an Op move, which suggests that at least one of the following cases must hold:

   i) either $\mathcal{DM}_j$ is replied to by $\mathcal{DM}_{i'} : Pr - C$ where $C \in Pr(\mathcal{T}')$ (given $\mathscr{P}_{CG}$, it may also be that $C = A$), which falls under the line of thinking employed in case II) steps c), d), e) and f), or;

ii) if i) is not the case, then $(B, A) \in \mathcal{C}^{\mathscr{D}}$ must be a preference-dependent attack, which particularly implies that $(B, A) \in \mathcal{C}^{\mathscr{D}}$ must be a preference-dependent attack on A', while additionally Pr must be able to undermine the success of B's attack on A as defeat.

b) a)ii) implies that Pr must have moved a preference-ordering $A' > B$ by means of a move:

$$\mathcal{DM}_{i'} : Pr - (A' > B)$$

against $\mathcal{DM}_j$.

c) Given b) we then differentiate between the following cases:

i) either Op was unable to introduce a counter preference-ordering $B > A'$ thus the dispute would have ended with $\mathcal{DM}_{i'}$, or;

ii) Op was able to introduce a counter preference-ordering $B > A'$ by means of a move:

$$\mathcal{DM}_{j'} : Op - (B > A')$$

against $\mathcal{DM}_{i'}$, in which case, due to $\mathscr{P}_{CG}$ and based on logical-completeness, Pr would have repeated $\mathcal{DM}_{i'}$ against $\mathcal{DM}_{j'}$ thus again ending the dispute with a Pr move.

d) If c)i) holds, then it must be that $(A' > B) \in \mathcal{P}^{\mathscr{D}}$, which by logical-completeness implies that $B \not> A'$ *does not hold* and thus, according to Definition 15, it must hold that $(B, A) \notin \mathcal{D}$, which in turn contradicts with condition 2.

$$d'_l : \boxed{\mathcal{DM}_r} \leftarrow \dots \leftarrow \boxed{\mathcal{DM}_i^{con} = A} \leftarrow \boxed{\mathcal{DM}_j^{con} = B} \leftarrow \boxed{\mathcal{DM}_{i'}^{con} = A' > B} \leftarrow \boxed{\mathcal{DM}_{j'}^{con} = B > A'} \leftarrow \boxed{\mathcal{DM}_{i'}}$$

Figure 3.9: Cases I)e)i) & I)e)ii)

e) If c)ii) holds, then out of the two contradictory preference-orderings moved into the game, $A' > B$ and $B > A'$, only the first would be included in $\mathcal{P}^{\mathscr{D}}$ (based on 37 for $\mathscr{P}_{CS}$), given that it was moved by Pr. Thus similarly, by logical-completeness $B \not> A'$ *does not hold*, and thus $(B, A) \notin \mathcal{D}$, which contradicts with condition 2.

Suppose that II) is the case. Then:

a) Given that the content of $\mathcal{DM}_i$ is an argument, there cannot exist dialogue moves in d'_l between $\mathcal{DM}_j$ and $\mathcal{DM}_i$ with preference-orderings as contents, since otherwise $\mathcal{DM}_i$ cannot follow.

b) a) implies that there must exist an argument $C \in d'_l$, and thus by condition of the winning strategy $C \in Pr(\mathcal{T}')$, such that $(C, B) \in \mathcal{C}^{\mathscr{D}}$ introduced against $\mathcal{DM}_j$ by Pr by means of a move:

$$\mathcal{DM}_{i'} : Pr - C$$

(where it may also be that $C = A$).

c) Suppose $(C, B) \in \mathcal{C}^{\mathscr{D}}$ is preference-independent. This immediately implies that $(C, B) \in \mathcal{D}$, which contradicts with condition 3.

d) Suppose $(C, B) \in \mathcal{C}^{\mathscr{D}}$ is preference-dependent and C attacks B on a sub-argument B'. Given that $C \in Pr(\mathcal{T}')$, then by condition of the winning strategy:

   i) either no preference-orderings related to B and C were introduced into the game by either Pr or Op, or;

   ii) if $\mathcal{DM}_{i'}$ was replied to by a move:

$$\mathcal{DM}_{j'} : Op - (B' > C)$$

   then $\mathcal{DM}_{j'}$ must have also been replied to by a counter preference-ordering $C > B'$ by means of a move:

$$\mathcal{DM}_{i''} : Pr - (C > B')$$

Note that otherwise, argument C cannot be part of d'_l and consequently of the winning strategy (advise Figures 3.10,3.11).

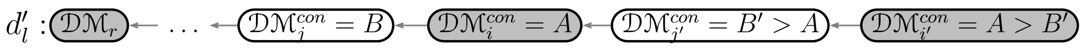


Figure 3.10: Case II)d)ii) for $C = A$

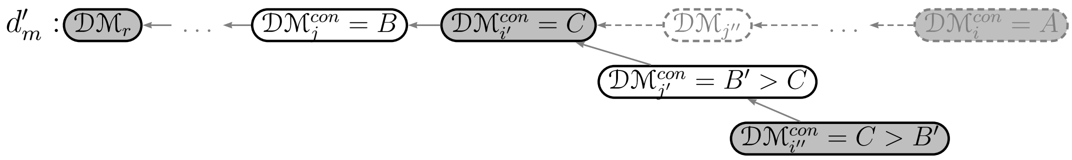


Figure 3.11: Case II)d)ii) for $C \neq A$

e) In case d)i), based on logical completeness it must hold that there exist no preference-orderings in $\mathcal{P}^{\mathscr{D}}$ with respect to C and B', which in turn according to Definition 15 implies that $C \not> B'$. The latter implies that $(C, B) \in \mathcal{D}$ which leads to a contradiction with condition 3.

f) If d)ii) holds, then out of the two contradictory preference-orderings introduced into the game $C > B'$ and $B' > C$, only the first would be included in $\mathcal{P}^{\mathscr{D}}$ based on Definition 37 for $\mathscr{P}_{CS}$, given that it was moved by Pr. Thus similarly, by logical-completeness it holds that $C \not> B'$, and thus $(C, B) \in \mathcal{D}$ which contradicts with condition 3.

In general, we have shown that the existence of an argument $B \in \mathcal{A}^{\mathscr{D}}$ able to invalidate the admissibility of $Pr(\mathcal{T}')$ through defeating an argument $A \in Pr(\mathcal{T}')$, leads to one of two possible contradictions; namely that either condition 2 or 3 does not hold.

Theorem 2 (Grounded Soundness) *Let $AF_{\mathscr{D}} = (A^{\mathscr{D}}, \mathcal{D})$ be the argumentation framework defined by a finite legal logically-complete grounded persuasion dialogue $\mathscr{D}$ for a claim c_{init}. Then if Pr is winning, there exists an argument X for c_{init} such that X is in the grounded extension of $AF_{\mathscr{D}}$.*

Proof Assume a finite legal logically-complete grounded persuasion dialogue $\mathscr{D}$ for a claim c_{init} where Pr is currently winning. Then based on Definition 35 and Lemma 1 there must exist a $\mathcal{T} \in \mathscr{F}$, the content of the root move of which is an argument X for c_{init}, with a winning strategy $\mathcal{T}' \in \mathcal{T}$ for Pr.

Let GE be the grounded extension in $AF_{\mathscr{D}}$. Then proving X's membership in GE can then be shown if $Pr(\mathcal{T}')$ is a particular subset of GE where:

- let $F_{AF_{\mathscr{D}}}$ be the characteristic function of $AF_{\mathscr{D}}$, as defined in Dung [1995], such that:
  - $F_{AF_{\mathscr{D}}} : 2^{\mathcal{A}^{\mathscr{D}}} \rightarrow 2^{\mathcal{A}^{\mathscr{D}}}$
  - $F_{AF_{\mathscr{D}}}(S) = \{A | A \text{ is acceptable with respect to } S\}$, for S a subset of $\mathcal{A}^{\mathscr{D}}$

  where $F_{AF_{\mathscr{D}}}$ is monotonic with respect to set inclusion. Furthermore:

- let $\sigma = S_0 \subseteq \ldots \subseteq S_n \subseteq \ldots$ be an increasing sequence of sets of arguments such that:
  - $S_0 = \varnothing$
  - each $S_i = F_{AF_{\mathscr{D}}}(S_{i-1})$

  while, as we always refer to an arbitrary but fixed AF, for convenience, we henceforth write $F^i(\varnothing)$ instead of $F^i_{AF_{\mathscr{D}}}(\varnothing)$, where the exponent, i, expresses the number of the iterative applications of F on $\varnothing$ for the production of the corresponding set S_i, then:

$$GE \subseteq \bigcup_{i=0}^{\infty} F^i(\varnothing) \tag{3.8}$$

We prove that $Pr(\mathcal{T}') \subseteq GE$, by induction.

We assume that a terminal move in a $\mathcal{T}$ (a leaf node) is referred to with an index τ, i.e. as $\mathcal{DM}_\tau$. Correspondingly, we will also refer to the level of a terminal node as τ, while we will refer to preceding levels as $\tau - i$, where $0 < i < \tau$.

Let $\{d'_1, \ldots, d'_m\}$ be the disputes in $\mathcal{T}'$, where each d'_i is of the form

$$d'_i = \{\mathcal{DM}_r, \ldots, \mathcal{DM}_{\tau-2}, \mathcal{DM}_{\tau-1}, \mathcal{DM}_\tau\}$$

and where by condition of the winning strategy $\mathcal{DM}_\tau$ is a proponent move. We distinguish between two kinds of disputes in a $\mathcal{T}'$ through categorising them in

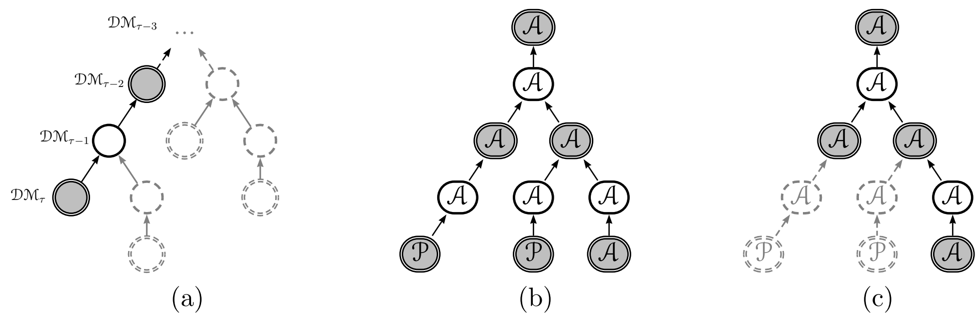


Figure 3.12: a) A terminal node $\mathcal{DM}_\tau$ in d' of a winning strategy $\mathcal{T}'$ and the moves that precede it in d', b) A winning strategy $\mathcal{T}'$ where the nodes marked with $\mathcal{A}$ and $\mathcal{P}$ contain arguments and preferences respectively, c) $\mathcal{T}'$ converted to $\mathcal{T}'^*$ through pruning off the preference nodes as well as the arguments they attack.

two distinct sets, $d\mathcal{P}$ and $d\mathcal{A}$, where $d\mathcal{P}$ contains those that end with a preference-ordering, and $d\mathcal{A}$ those that end with an argument, such that:

$$\mathcal{T}' = d\mathcal{P} \cup d\mathcal{A}$$

For building our inductive hypothesis, we rely on the following observations:

Observation 1 *There do not exist any preference-orderings moved by Op in a $\mathcal{T}'$.*

Observation 2 *Every argument moved by Pr in a $\mathcal{T}'$ is a defeating argument.*

Observation 3 *All preference-orderings moved in $\mathcal{T}'$ appear only in leaf nodes.*

The first observation results from the fact that if Op introduces a preference-ordering $X > Y$ against a Pr attack in $d' \in \mathcal{T}'$ then, based on the $\mathscr{P}_{GG}$ protocol rules, if Pr is aware of a counter preference-ordering $Y > X$, its introduction is useless as Op is able to repeat $X > Y$. The latter means that d' will end with an Op move which contradicts with the winning strategy definition.

Observation 2 essentially follows form the first observation, as, since Op cannot introduce any preference-orderings in a d', it is evident that every argument moved by Pr will only be attacked by other Op arguments, which implies that all Pr attacks succeed as defeats.

Finally, the third observation is a combination of the fact that preference-orderings can only be attacked by counter preference-orderings, and of the winning strategy definition. Suppose for example that Pr moves a preference-ordering $X > Y$ in a non-leaf node of a d', then Op's only possible response would be the introduction of a counter preference-ordering $Y > X$ which, by condition of $\mathscr{P}_{GG}$, implies that the dispute will end with an Op move. The latter obviously contradicts with the winning strategy definition. Therefore, in the case of a $\mathcal{T}'$ preference-orderings only appear in leaf nodes and are only moved by Pr.

From the above observations one may conclude that, in contrast with Pr, some of Op's attacks in a $\mathcal{T}'$ are unsuccessful, i.e. those found in $\tau - 1$ nodes which are attacked by preference-orderings moved by Pr. In this respect, if we prune off all the leaf nodes in a $\mathcal{T}'$ that contain Pr preferences, as well as the $\tau - 1$ nodes they attack, we end up with a sub-tree $\mathcal{T}^* = \{d_1^*, d_2^*, \ldots, d_m^*\}$ of $\mathcal{T}'$ which only contains defeats (see Figure 3.12c). In essence, let $\mathscr{D}_a$ and $\mathscr{D}_b$ be two persuasion dialogues on the acceptability of a claim x, and $\mathcal{T}_a$ and $\mathcal{T}_b$ respectively the corresponding dialogue trees, where the participants exchange:

a) arguments and preferences based on their attack relation, resulting in the instantiation of a $PAF =< \mathcal{A}^{\mathscr{D}}, \mathcal{C}^{\mathscr{D}}, \mathcal{P}^{\mathscr{D}} >$ as defined in Definition 37, where the conflicts between preferences are resolved for $\mathcal{S} = Grounded$ and;

b) just arguments based on their defeat relation, relying on a framework $AF_{\mathscr{D}} =< \mathcal{A}^{\mathscr{D}}, \mathcal{D} >$ instantiated from the induced PAF in case a)

Then it must hold that:

$$\exists \mathcal{T}'_a \in \mathcal{T} \Leftrightarrow \exists \mathcal{T}^*_b \in \mathcal{T}_b$$

in which case $\mathcal{T}'_a = \mathcal{T}^*_b$. In this sense, proving that $Pr(\mathcal{T}') \subseteq GE$ is equivalent with proving that $Pr(\mathcal{T}^*) \subseteq GE$.

- **Basis:** Based on observations 1,2 & 3, it is evident that all the Pr arguments found in the leaf nodes of a $\mathcal{T}^*$ are undefeated. In this respect, let $\texttt{term}(d_i^*)$ be a function such that:

$$\texttt{term}(d_i^*) = \texttt{Content}(\mathcal{DM}_\tau \in d_i^*)$$

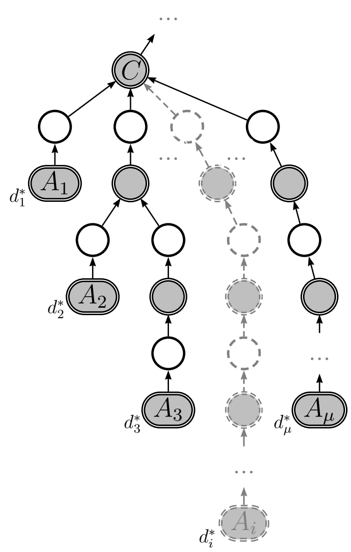


Figure 3.13: An argument C in a $\mathfrak{T}^*$ whose acceptability is contingent upon whether it is distinctly acceptable in all the disputes $\{d_1^*, d_2^*, d_3^*, \ldots, d_i^*, \ldots, d_\mu^*\}$ it participates, for $\mu \leq m$.

Hence, since by definition of the fixed point function, the application of $F^1(\varnothing)$ results in the collection of all undefeated arguments in an $AF_{\mathscr{D}}$, it then holds that:

$$W = \bigcup_{i=1}^{m} \texttt{term}(d_i^*) \subseteq F^1(\varnothing) \tag{3.9}$$

- **Step:** Based on equations 3.8 & 3.9 it is evident that, since W is a subset of $F^1(\varnothing)$, the iterative application of the F on W will produce a subset of the GE, i.e.:

$$\bigcup_{i=0}^{\infty} F^i(W) \subseteq GE \tag{3.10}$$

Let:

$$\sigma_W = W \subseteq F^1(W) \subseteq F^2(W) \subseteq \ldots \subseteq F^i(W) \subseteq \ldots$$

then, for proving that $Pr(\mathfrak{T}') \subseteq GE$ it is enough that we prove that every element of $Pr(\mathfrak{T}')$ is an element of some $F^i(W)$.

A naive approach would be to assume that since W contains all the undefeated arguments found in the leaf nodes of a $\mathfrak{T}^*$, then $F^1(W)$ can be produced by the union product of W and of all the $\tau - 2$ level arguments in $\mathcal{T}^*$, as, intuitively, those should be supported by the elements of W (undefeated arguments). Nevertheless, this does not hold as it might be the case that a $\tau - 2$ level argument in a certain dispute d_i^*, is also a $\tau - k$ level argument in another dispute d_j^* in $\mathcal{T}^*$, where $k \in 2\mathbb{N}$ and $k > 2$. For example, argument C in Figure 3.13 participates in μ disputes where it is a $\tau - 2$ level argument in d_1^*, a $\tau - 4$ in d_2^*, $\tau - 6$ in d_3^* etc. As C participates in more than one disputes with varying depths, it is not possible for it to be deemed acceptable with respect to just W (particularly with respect to just argument $A_1 \in W$ the latter being an element of d_1^*) and therefore be an element of $F^1(W)$, as not all the attacks on C are accounted by the elements of W.

Hence, $C \in Pr(\mathcal{T}^*)$ can only be an element of some $F^i(W)$ if it is distinctly acceptable in all the disputes it participates. For proving C's acceptability we define a fixed point function $f_{d_i^*}$ which is similar to F but which is distinctly applied on each of the set of arguments $\mathcal{A}_{d_i^*} \subseteq \mathcal{A}^{\mathscr{D}}$ found on each of the disputes in a $\mathcal{T}^*$, defined as follows:

Let $\mathcal{T}^*$ be a pruned winning strategy such that $\mathcal{T}^* = \{d_1^*, d_2^*, \ldots, d_m^*\}$, then $f_{d_i^*}$ is the characteristic function of $AF_{d_i^*} \subseteq AF_{\mathscr{D}}$, where $AF_{d_i^*} = \{\mathcal{A}_{d_i^*} \subseteq \mathcal{A}^{\mathscr{D}}, \mathcal{D}_{d_i^*} \subseteq \mathcal{D}\}$, such that:

- $f_{d_i^*} : 2^{\mathcal{A}_{d_i^*}} \rightarrow 2^{\mathcal{A}_{d_i^*}}$
- $f_{d_i^*}(S) = \{A | A \text{ is acceptable with respect to } S\}$, for S a subset of $\mathcal{A}_{d_i^*}$

where $f_{d_i^*}$ is monotonic with respect to set inclusion. In a similar sense to F, assuming an $f_{d_i^*}^k$, the exponent k expresses the number of the iterative applications of f on $\varnothing$.

In essence, the acceptable set of arguments produced by the iterative application of f, for k times, on the set of arguments found in some d_i^* is:

$$f_{d_i^*}^k(\varnothing) = \bigcup_{j=0}^{k} \texttt{Content}(\mathcal{DM}_{\tau-2j}) \tag{3.11}$$

In this sense, and based on equations 3.11, 3.12 and 3.11, it easily follows that:

$$W = \bigcup_{i=0}^{m} f_{d_i^*}(\varnothing) \tag{3.12}$$

Let $\texttt{depth}()$ be a function that takes as input a tree and returns its depth. Then the number of times that the characteristic function F can be applied on a set of arguments found in a dialogue tree $\mathcal{T}$, which essentially represent the set $\mathcal{A}^{\mathscr{D}}$, until a fixed point is reached, is:

$$n_{\mathcal{T}} = \frac{\texttt{depth}(\mathcal{T}) - 1}{2}$$

This is the number of times, at least necessary, for accounting for all the defeats against the root argument of $\mathcal{T}^*$, and which is equal to the depth of the longest dispute d_i^* of $\mathcal{T}^*$, divided by two—as every argument in a d_i^* can only be acceptable with respect to an argument 2 levels after it in d_i^*. Thus, the fixed point of F applied on W is reached after $n_{\mathcal{T}^*}$ steps.

Equivalently, construction of $F^{n_{\mathcal{T}^*}}(W)$ can be achieved through the union product of all the fixed points $\{f_{d_1^*}^{n_{d_1^*}}(\varnothing), f_{d_2^*}^{n_{d_2^*}}(\varnothing), \ldots, f_{d_m^*}^{n_{d_m^*}}(\varnothing)\}$ of $f_{d_1^*}(\varnothing)$, $f_{d_2^*}(\varnothing), \ldots, f_{d_m^*}(\varnothing)$ of all the disputes (m) of a pruned winning strategy $\mathcal{T}^*$. This holds, as each defeat on the root argument of $\mathcal{T}^*$ is accounted by the fixed point of f applied on that defeat's corresponding dispute, i.e. since the root argument is acceptable with respect to the fixed points of f applied on the arguments found on each of the disputes of $\mathcal{T}^*$. Essentially, since due to logical completeness if there exists an argument that could have been put forth against another argument in $Pr(\mathcal{T}')$ it has been put forth, then the existence of a winning strategy $\mathcal{T}^*$ captures all possible defeats against an argument in $Pr(\mathcal{T}^*)$ each of which are accounted by the iterative application of f on each of the disputes of $\mathcal{T}^*$.

We can therefore construct the fixed point of F applied on W, as follows:

$$F^{n_{\mathcal{T}^*}}(W) = \bigcup_{i=0}^{m} f_{d_i^*}^{n_{d_i^*}}(\varnothing) \tag{3.13}$$

Finally, let C be an argument in $Pr(\mathcal{T}^*)$, and $\mathcal{T}_C = \{d_1^C, d_2^C, \ldots, d_\mu^C\}$ a sub-tree of $\mathcal{T}^*$ with C as its root node. Then W_C is a subset of W where:

$$W_C = \bigcup_{i=0}^{\mu} f_{d_i^C}(\varnothing) \tag{3.14}$$

it then holds that:

$$F^{n_{\mathcal{T}_C}}(W_C) = \bigcup_{i=0}^{\mu} f_{d_i^C}^{n_{d_i^C}}(\varnothing) \tag{3.15}$$

Given that each d_i^C is a sub-dispute of some d_j^* in $\mathcal{T}^*$, and since $F^{n_{\mathcal{T}_C}}(W_C)$, based on Equation 3.15 and due to logical completeness, accounts for all the defeats on C, it is then evident that:

$$F^{n_{\mathcal{T}_C}}(W_C) \subseteq F^{n_{\mathcal{T}_C}}(W)$$

Since C is an element of $F^{n_{\mathcal{T}_C}}(W)$ it is also an element of GE.

Theorem 3 (Credulous Fairness) *Given a finite $AF_{\mathscr{D}} = (\mathcal{A}^{\mathscr{D}}, \mathcal{D})$, where S is in a admissible extension of $AF_{\mathscr{D}}$, then for any $X \in S$ there exists a finite logically-complete legal credulous persuasion dialogue $\mathscr{D}$, such that Pr is winning.*

Proof Let $AF_{\mathscr{D}} = (\mathcal{A}^{\mathscr{D}}, \mathcal{D})$ be an argumentation framework induced by a $PAF = < \mathcal{A}^{\mathscr{D}}, \mathcal{P}^{\mathscr{D}}, \mathcal{C}^{\mathscr{D}} >$, where $S \subseteq \mathcal{A}^{\mathscr{D}}$ is an admissible extension in $AF_{\mathscr{D}}$, $|S|$ the number of elements in S, and $\mathcal{Q} = \mathcal{A}^{\mathscr{D}} - S$. Assume then that for an argument $X \in S$ a legal, logically complete, credulous persuasion dialogue game will take place.

Let us assume that the game is initiated by Pr's move such that $d_0 : Pr - \mathcal{DM}_{i=0}$, where $\mathcal{DM}_{i=0}^{con} = X$. Given that S is an admissible set where $|S| = n$, then $\forall X \in S$ one of the following must hold:

A) $\nexists Y \in \mathcal{Q}$ such that $(Y, X) \in \mathcal{C}^{\mathscr{D}}$ (there is no argument that attacks X), or;

B) $\forall Y \in \mathcal{Q}$ where $(Y, X) \in \mathcal{C}^{\mathscr{D}}$, then:

- I) either $(X, Y) \in \mathcal{C}^{\mathscr{D}}$, $(X, Y) \in \mathcal{D}$, $(Y, X) \notin \mathcal{D}$ and Pr only uses X, or;
- II) $(X, Y) \notin \mathcal{C}^{\mathscr{D}}$, $(Y, X) \notin \mathcal{D}$, where Pr only uses preferences or;

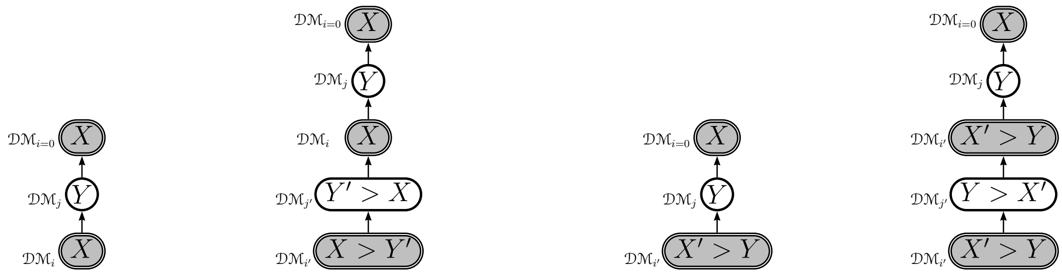
C) let $S_{\mathcal{D}} \subseteq \mathcal{D}$ such that:

$$S_{\mathcal{D}} = \{(Y, X) : Y \in \mathcal{Q}, (Y, X) \in \mathcal{D}\}$$

then there exists a set $S'_{\mathcal{D}}$, where $S'_{\mathcal{D}} \subseteq \mathcal{D}$ and $S'_{\mathcal{D}} \cap S_{\mathcal{D}} = \varnothing$, such that:

$$S'_{\mathcal{D}} = \{(Z, Y) : \forall (Y, X) \in S_{\mathcal{D}}\ \exists (Z, Y) : Z \in S, Y \in \mathcal{Q}, (Z, Y) \in \mathcal{D}\}$$

i.e., for every Y that defeats an X in S there exists a Z in S that defeats Y.



(a) (b) (c) (d)

Figure 3.14: Credulous fairness: Cace B): Possible sub-cases: a) Case B)I)a), b) Case B)I)b), c) Case B)II)a), d) Case B)II)b)

Our goal is to show that for each of these cases, the possible dialogues that may be induced with a root move $\mathcal{DM}_0^{con} = X$, will definitely be a *winning strategy* for Pr and that therefore, Pr will always be the winner. To do so we rely on showing that each of the possible disputes that begin with X will end with a Pr move.

• **Let A) be the case, then:**

The dialogue will trivially terminate with $d_0 : Pr - X$. Thus $d_0 \in \mathcal{T}'$ which makes Pr the winner.

• **Let B)I) be the case, then:**

By completeness conditions and due to $\mathscr{P}_{CG}$, according to which Op cannot repeat a move in the same dispute while Pr can, the dialogue will be extended with a move $\mathcal{DM}_j^{con} = Y$, while $\mathcal{DM}_i$ will be repeated against $\mathcal{DM}_j$:

$$d_0, Op - Y, Pr - X$$

We further differentiate between the following cases:

a) either $(X, Y) \in \mathcal{C}^{\mathscr{D}}$ is a preference-independent attack, or;

b) $(X, Y) \in \mathcal{C}^{\mathscr{D}}$ is a preference-dependent attack on Y'.

∘ **If a) is the case:**

i) This directly implies that $(X, Y) \in \mathcal{D}$.

ii) In addition, based on $\mathscr{P}_{CG}$—according to which Op cannot repeat the same move in the same dispute line—Y cannot be repeated against X.

iii) Based on i) and ii) the dispute will end with Pr's move (Figure 3.14a).

∘ **If b) is the case, then:**

i) If Op is not aware of a preference-ordering $Y' > X$ able to undermine the success of $(X, Y) \in \mathcal{C}^{\mathscr{D}}$ as defeat, then, similarly to B)I)a), the dialogue will end with Pr's move (Figure 3.14a).

ii) Else, if Op is aware of a preference-ordering $Y' > X$ this would be moved as the content of a move $\mathcal{DM}_{j'}$ against the repeated $\mathcal{DM}_i$:

$$d_0, Op - Y, Pr - X, Op - (Y' > X)$$

1. Since by assumption $(X, Y) \in \mathcal{D}$, Pr must also be aware of a counter preference-ordering $X > Y'$, which will be moved as the content of a move $\mathcal{DM}_{i'}$ against $\mathcal{DM}_{j'}$ (advise Figure 3.14b):

$$d_0, Op - Y, Pr - X, Op - (Y' > X), Pr - (X > Y')$$

2. Based on $\mathscr{P}_{CG}$, which dictates that Op cannot repeat its moves in the same dispute, the dispute will again end with Pr's move.

3. Notice that, by Definition 37 for the credulous game, we keep $X > Y' \in \mathcal{P}^{\mathscr{D}}$ at the expense of $Y' > X$.

4. Note that as we assume the concerned dialogue to be taking place on the basis of a PAF, and given that set $\mathcal{P}^{\mathscr{D}}$ is consistent by definition, move $\mathcal{DM}_{j'}$ can never appear in the concerned dispute.

5. Thus, the dispute will actually end with the repetition of $\mathcal{DM}_i$ by Pr (Figure 3.14a).

• **Let B)II) be the case, then:**

By completeness conditions the dialogue will be extended with a move $\mathcal{DM}_j$: $Op{-}Y$. Since by assumption $(Y, X) \notin \mathcal{D}$, then $(Y, X) \in \mathcal{C}^{\mathscr{D}}$ must be a preference-dependent attack on X', while Pr must be aware of a preference-ordering $X' > Y$ which will be introduced into the game against $\mathcal{DM}_j$ as the content of a move $\mathcal{DM}_{i'}$:

$$d_0, Op - Y, Pr - (X' > Y')$$

We then differentiate between the following cases:

a) Either Op is *not* aware of a counter preference-ordering $Y > X'$, or;

b) Op is aware of a counter preference-ordering $Y > X'$.

◦ **If a) is the case, then:**

i) $\mathcal{DM}_i$ will simply be the last move in the dispute (Figure 3.14c).

◦ **If b) is the case, then:**

i) $Y > X'$ will be moved into the game as the content of a move $\mathcal{DM}_{j'}$, against $\mathcal{DM}_{i'}$:

$$d_0, Op - Y, Pr - (X' > Y), Op - (Y > X')$$

ii) By $\mathscr{P}_{CG}$, Pr can repeat its move—$\mathcal{DM}_{i'}$—against $\mathcal{DM}_{j'}$ in the same dispute, in contrast with Op:

$$d_0, Op - Y, Pr - (X' > Y), Op - (Y > X'), Pr - (X' > Y)$$

1. Thus the dispute will end again with Pr's move (Figure 3.14d).

2. Notice again that, by Definition 37 for the credulous game, we keep $X' > Y \in \mathcal{P}^{\mathcal{D}}$ at the expense of $Y > X'$.
3. As we assume the concerned dialogue to be taking place on the basis of a PAF and given that set $\mathcal{P}^{\mathcal{D}}$ is consistent by definition, move $\mathcal{DM}_{j'}$ can never appear in the concerned dispute.
4. Thus, the dispute will actually end with move $\mathcal{DM}_{i'}$ by Pr (Figure 3.14c).

Therefore in case B) all the possibly induced disputes can end either with the repetition of X or with a preference-ordering $X' > Y_i$ for $i = 0, \ldots, k$ (Figure 3.15).

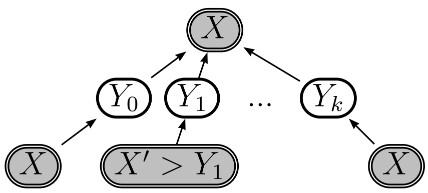


Figure 3.15: A possible dialogue tree induced for argument X and all the possible Ys that may be moved against it, which fall into cases A) and B).

- **Let C) be the case then:**

  Suppose S is defined as follows:

$$S = \{X_1, X_2, \ldots, X_{I-1}, X_I, X_{I+1}, \ldots, X_n\}$$

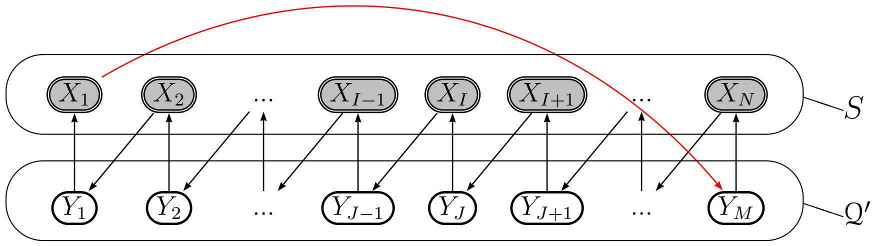
while we assume a set $\mathcal{Q}' \subseteq \mathcal{Q}$ where:

$$\mathcal{Q}' = \{Y_1, Y_2, \ldots, Y_{J-1}, Y_J, Y_{J+1}, \ldots, Y_m\}$$

where every element in $\mathcal{Q}'$ is a defeater of some element in S.

Suppose Pr initiates a game for an argument X_I moved into the game as the content of a move $\mathcal{DM}_{I=0}$, while there exists a Y_J such that $(Y_J, X_I) \in S_{\mathcal{D}}$ and $(X_I, Y_J) \notin \mathcal{D}$. Let $mod(a, b)$ be a function that returns the remainder of division of a by b. Then:

i) The dispute will be extended with Y_J being moved against X_I as the content of a move $\mathcal{DM}_j$.



(a)

(b)

Figure 3.16: a) Sets S and $\mathcal{Q}'$, and a possible defeat relationship between their elements, b) A dispute where the initiating argument is X_1

ii) As S is an admissible set, assume that it holds that $(X_{mod(I+1,N)}, Y_J) \in S'_{\mathcal{D}}$.

iii) The dispute will be further extended with $X_{mod(I+1,N)}$ being moved against Y_J as the content of a move $\mathcal{DM}_{i'}$.

iv) Assume then that there exists a $Y_{mod(J+1,M)}$ such that:

$$(Y_{mod(J+1,M)}, X_{mod(I+1,N)}) \in S_{\mathcal{D}} \;\; and \;\; (X_{mod(I+1,N)}, Y_{mod(J+1,M)}) \notin \mathcal{D}$$

i.e., the attacked argument does not counter-attack the attacker.

v) The dispute will be extended with $Y_{mod(J+1,M)}$ being moved against $X_{mod(I+1,N)}$ as the content of a move $\mathcal{DM}_{j'}$.

vi) As S is an admissible set, then for every $Y_{mod(I+k+1,M)}$, where $k = 1, 2, \ldots$, for which for some $X_{mod(I+k,N)}$ holds that:

$$(Y_{mod(J+k+1,M)}, X_{mod(I+k,N)}) \in S_{\mathcal{D}}$$

and which is moved into the dispute by Op, we can assume that $\exists X_{mod(I+k+1,M)}$ such that:

$$(X_{mod(I+k+1,M)}, Y_{mod(J+k+1,M)}) \in S'_{\mathcal{D}}$$

through which Pr will further extend the dispute moving against Op's move.

vii) As both S and $\mathcal{Q}'$ are finite sets (Figure 3.16a), in a worst case scenario for some $X_{mod(I+k+1,M)}$ where:

$$(X_{mod(I+k+1,M)}, Y_{mod(J+k+1,M)}) \in S'_{\mathcal{D}}$$

should hold that

$$mod(I+k+1, M) = I$$

viii) Thus X_I will be repeated in to the dispute by Pr.

ix) Nevertheless, while $\exists(Y_J, X_I) \in S_{\mathcal{D}}$, according to $\mathscr{P}_{CG}$ argument Y_J cannot be repeated into the dispute by Op.

x) Thus the dispute will terminate with Pr's move (Figure 3.16b).

Generally, in all three cases, A), B) and C), all the possible disputes that can be induced from $AF_{\mathscr{D}}$ with a root move $\mathcal{DM}_0^{con} = X$, where $X \in S$, and S is an admissible extension, can only end with a Pr move. This implies that any produced dialogue will also be a *winning strategy* for Pr. Therefore Pr will always be the winner.

Theorem 4 (Grounded Fairness) *Given a finite $AF_{\mathscr{D}} = (\mathcal{A}^{\mathscr{D}}, \mathcal{D})$, and S is in the grounded extension of $AF_{\mathscr{D}}$, then for any $X \in S$ there exists a finite logically-complete legal grounded persuasion dialogue $\mathscr{D}$, such that Pr is winning.*

Proof Let $AF_{\mathscr{D}} = (\mathcal{A}^{\mathscr{D}}, \mathcal{D})$ be an argumentation framework induced by a $PAF =< \mathcal{A}^{\mathscr{D}}, \mathcal{P}^{\mathscr{D}}, \mathcal{C}^{\mathscr{D}} >$, where $GE \subseteq \mathcal{A}^{\mathscr{D}}$ is the grounded extension in $AF_{\mathscr{D}}$, $|GE|$ the number of elements in GE, and $\mathcal{Q} = \mathcal{A}^{\mathscr{D}} - GE$. Assume then that for an argument $X \in GE$ a legal, logically complete, grounded persuasion dialogue game will take place.

Let us assume that the game is initiated by Pr's move such that $d_0 =< Pr, \mathcal{DM}_{i=0} >$, where $\mathcal{DM}_{i=0}^{con} = X$. Given that GE is a set where $|GE| = n$, then $\forall X \in GE$ one of the following must hold:

A) $\nexists Y \in \mathcal{Q}$ such that $(Y, X) \in \mathcal{C}^{\mathscr{D}}$ (there is no argument that attacks X), or;

B) $\forall Y \in \mathcal{Q}$ where $(Y, X) \in \mathcal{C}^{\mathscr{D}}$, then $(Y, X) \notin \mathcal{D}$ (Y does not defeat X), or;

C) let $S_{\mathcal{D}} \subseteq \mathcal{D}$ such that:

$$S_{\mathcal{D}} = \{(Y, X) : Y \in \mathcal{Q}, (Y, X) \in \mathcal{D}\}$$

then there exists a set $S'_{\mathcal{D}}$, where $S'_{\mathcal{D}} \subseteq \mathcal{D}$ and $S'_{\mathcal{D}} \cap S_{\mathcal{D}} = \varnothing$, such that:

$$S'_{\mathcal{D}} = \{(Z, Y) : \forall (Y, X) \in S_{\mathcal{D}}\ \exists (Z, Y) : Z \in GE, Y \in \mathcal{Q}, (Z, Y) \in \mathcal{D}, Z \neq X\}$$

while

$$\exists (Z, Y) \in S'_{\mathcal{D}} : \nexists (W, Z) \in S_{\mathcal{D}}$$

i.e., for every Y that defeats an X in GE there exists a Z in GE that defeats Y, while there exists at least one such Z for which there exists no defeater in $\mathcal{Q}$.

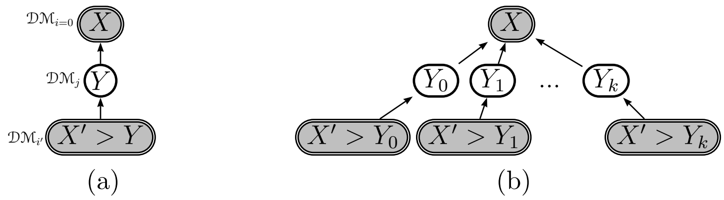


Figure 3.17: a) Case B), b) A possible dialogue tree induced for argument X and all the possible Ys that may be moved against it, which though also fall into Case B).

Our goal is to show that for each of these cases, the possible dialogues that may be induced with a root move $\mathcal{DM}_0^{con} = X$, will definitely be a *winning strategy* for Pr and that therefore, Pr will always be the winner. To do so we rely on showing that each of the possible disputes that begin with X will end with a Pr move.

• **Let A) be the case then:**

The dialogue will trivially terminate with $d_0 : Pr - X$. Thus $d_0 \in \mathcal{T}'$ which makes Pr the winner.

• **Let B) be the case then:**

i) The dialogue will be extended with a move $\mathcal{DM}_j^{con} = Y$:

$$d_0.Op - Y$$

ii) Due to $\mathscr{P}_{GG}$, according to which Pr cannot repeat a move in the same dispute while Op can, and since by assumption $(Y, X) \notin \mathcal{D}$ then:

1. In the case where $(X, Y) \in \mathcal{C}^{\mathscr{D}}$, $\mathcal{DM}_i$ will not be repeated against $\mathcal{DM}_j$.
2. $(Y, X) \in \mathcal{C}^{\mathscr{D}}$ must be a preference-dependent attack on X'.
3. Pr must be aware of a preference-ordering $X' > Y$ which will be moved into the game as the content of a move $\mathcal{DM}_{i'}$ against $\mathcal{DM}_j$ (Figure 3.17a):

$$d_0.Op - Y, Pr - (X' > Y)$$

4. Notice that since the dialogue is taking place on the basis of a PAF and given that set $\mathcal{P}^{\mathscr{D}}$ is consistent by definition, Op cannot be aware of a counter-preference ordering $Y > X'$, since otherwise by Definition 37 for the grounded game, we would keep:

$$Y > X' \in \mathcal{P}^{\mathscr{D}}$$

at the expense of:

$$X' > Y$$

5. Thus the dispute will end with Pr's move.

Therefore in case B) all possible disputes can only end with a preference-ordering $X' > Y_i$ for $i = 0, \ldots, k$, for every Y that attacks X (Figure 3.17b).

• **Let C) be the case then:**

Suppose S is defined as follows:

$$S = \{X_1, X_2, \ldots, X_{I-1}, X_I, X_{I+1}, \ldots, X_N\}$$

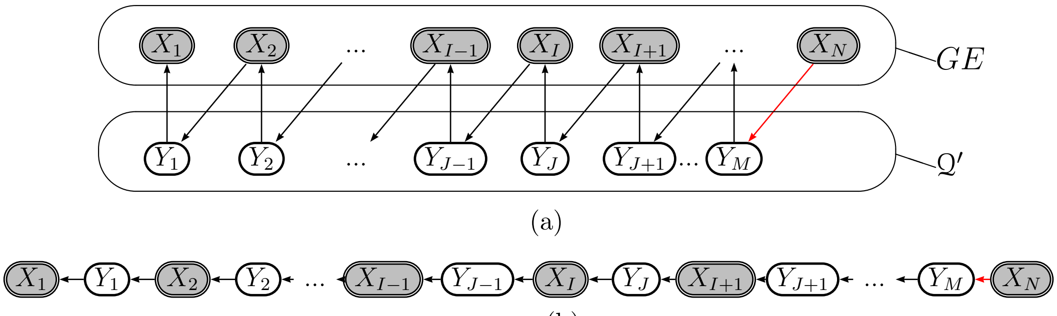


Figure 3.18: a) Sets S and $\mathcal{Q}'$, and a possible defeat relationship between their elements, b) A dispute where the initiating argument is X_1

while we assume a set $\mathcal{Q}' \subseteq \mathcal{Q}$ where:

$$\mathcal{Q}' = \{Y_1, Y_2, \dots, Y_{J-1}, Y_J, Y_{J+1}, \dots, Y_M\}$$

where every element in $\mathcal{Q}'$ is a defeater of some element in S.

Suppose Pr initiates a game for an argument X_I moved into the game as the content of a move $\mathcal{DM}_{i=0}$, while there exists a Y_J such that $(Y_J, X_I) \in S_{\mathcal{D}}$ and $(X_I, Y_J) \notin \mathcal{D}$. Then:

i) The dispute will be extended with Y_J being moved against X_I as the content of a move $\mathcal{DM}_j$.

ii) By assumption it must hold that $(X_{I+1}, Y_J) \in S'_{\mathcal{D}}$.

iii) The dispute will be further extended with X_{I+1} being moved against Y_J as the content of a move $\mathcal{DM}_{i'}$.

iv) Assume then that there exists a Y_{J+1} such that:

$$(Y_{J+1}, X_{I+1}) \in S_{\mathcal{D}} \ \ \textit{and} \ \ (X_{I+1}, Y_{J+1}) \notin \mathcal{D}$$

v) The dispute will be extended with Y_{J+1} being moved against X_{I+1} as the content of a move $\mathcal{DM}_{j'}$.

vi) As GE is the grounded extension, then for every Y_{I+k+1}, where $k = 1, 2, \ldots$, for which for some X_{I+k} holds that:

$$(Y_{J+k+1}, X_{I+k}) \in S_{\mathcal{D}}$$

which is moved into the dispute by Op, we can assume that $\exists X_{I+k+1}$ such that:

$$(X_{I+k+1}, Y_{J+k+1}) \in S'_{\mathcal{D}}$$

through which Pr will further extend the dispute moving it against Op's move.

vii) As both GE and $\mathcal{Q}'$ are finite sets and since GE is the grounded extension, after exhaustively using the elements of GE, there will be an element $X_{I+k+1} \in GE$ (which in a worst case scenario will be the *last* element in GE), where:

$$(X_{I+k+1}, Y_{J+k+1}) \in S'_{\mathcal{D}}$$

and for which it should hold that: $\nexists (Y_{J+k+2}, X_{I+k+1}) \in S_{\mathcal{D}}$

viii) Thus, the content of the of the dispute's last move will be X_{I+k+1} moved by Pr.

Generally, in all three cases, A), B) and C), all the possible disputes that can be induced from $AF_{\mathscr{D}}$ with a root move $\mathcal{DM}_0^{con} = X$, where $X \in GE$, and GE is the grounded extension, can only end with a Pr move. This implies that any produced dialogue will also be a *winning strategy* for Pr. Therefore Pr will always be the winner.

3.4 Strategic Considerations

In this section we deal with the possible strategic considerations that may be employed by a participant in a persuasion dialogue game. Particularly we focus on illustrating how not taking into account the underlying logic fails again in accounting for the dynamic nature of a dialogue game. The basic idea is to illustrate particularly, how accounting for the underlying logic may have a considerable effect on the outcome of a dialogue game. In order to do so, we need to first show how an abstract approach differs from one that accounts for the underlying logic. For this purpose, we rely on the following example which aims to stress the impact of backtracking on strategy planning.

3.4.1 The Importance of Backtracking

We begin by illustrating how backtracking may affect the outcome of a dialogue game, through comparing two protocols for the grounded semantics, where the first does not allow for it while the second does.

Example 1 *Let there be a persuasion dialogue $\mathscr{D}$ where:*

- *the goals of the participants, Pr and Op, concern proving ($\mathcal{G}_{Pr}$) respectively disproving ($\mathcal{G}_{Op}$) the acceptability of an argument X in dispute, with respect to the grounded semantics*

- *the participants exchange arguments based on a binary attack relationship $\mathcal{C}$, Pr begins with X*

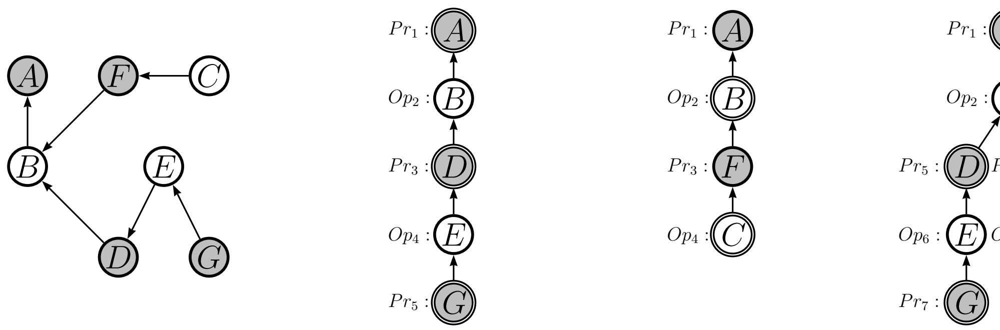
- *we assume two protocols $\mathscr{P}_{GS1}$ and $\mathscr{P}_{GS2}$ where backtracking is not respectively is allowed*

- *assuming $\mathscr{D}$ is expressed in the form of a tree $\mathcal{T} = \{d_1, ...d_L\}$ with X as its root, then given a labelling applied on $\mathcal{T}$ based on Definition 33, an argument can be introduced in $\mathscr{D}$ only if it changes the dialogical status of X*

- *both $\mathscr{P}_{GS1}$ and $\mathscr{P}_{GS2}$ dictate that $\forall d \in \mathcal{T}$, no two arguments with speaker Pr are the same, and; the game terminates if one of the players cannot respond in any way*

Let us then assume, that in order for the participants to strategise, they rely on building a similar opponent model as the one described in Section 3.1, with respect though only to the arguments and their binary attack relationship, assumed to be known by their opponents.

We are interested in investigating the possible ways based on which a game on the acceptability of an argument A may evolve, based on the perspective of the proponent. Therefore, let us further assume that Pr: is aware of arguments



(a) (b) (c) (d)

Figure 3.19: a) An AF instantiated based on both an agents own KB and its opponent's model, b) Scenario I:[$\mathscr{P}_{GS1}$] [$\mathscr{P}_{GS2}$], c) Scenario II:[$\mathscr{P}_{GS1}$], d) Scenario II:[$\mathscr{P}_{GS2}$], the simulated dialogue tree.

$\{A, D, F, G\}$; assumes that Op believes $\{B, C, E\}$. Given the binary attack relationships between the elements of these sets of arguments, we assume that for the purpose of strategising, Pr instantiates a new abstract AF (Figure 3.19a) combining his own beliefs (grey nodes) with his assumptions about the beliefs of Op (white nodes). For the purpose of this example we assume that Pr's beliefs about the beliefs of Op are `Complete`.

Based on the AF depicted in Figure 3.19a[1], Pr, who is to prove the acceptability of A, is able to simulate the possible ways based on which the dialogue may evolve, with respect to the different protocols, as described by the following scenarios:

Scenario I:[$\mathscr{P}_{GS1}$] *Pr moves A. Op responds with B. Out of the two possible options Pr responds with D. Consequently, Op follows with E and finally Pr ends the game with G, since Op has no way to counter G (Figure 3.19b).*

Scenario II:[$\mathscr{P}_{GS1}$] *Pr moves A. Op responds with B. Out of the two possible options Pr responds with F. Consequently, Op ends the game with C, since*

[1]In Figures 3.19b, 3.19c, 3.19d, the arrows depicted are deem, since they do not represent actual attack relationships between dialogue moves, but rather concern the sequence order in which they are introduced

according to the game's rules Pr is not allowed to backtrack (Figures 3.19c).

Scenario I:[$\mathscr{P}_{GS2}$] *Pr moves A. Op responds with B. Out of the two possible options Pr responds with D. Consequently, Op follows with E and finally Pr ends the game with G, since Op has no way to counter G or any option for backtracking, even though backtracking is allowed (Figure 3.19b).*

Scenario II:[$\mathscr{P}_{GS2}$] *Pr moves A. Op responds with B. Out of the two possible options Pr responds with F. Op responds with C. Since there is no way of countering C, Pr now can only backtrack to countering B with D. Op follows with E and finally Pr ends the game with G, since Op has no way to counter G or any option for backtracking (Figure 3.19d).*

Based on this example, one is able to observe that in $\mathscr{P}_{GS2}$, Pr is able to reach the desirable outcome regardless of the choices it makes through the game. The latter does not hold for $\mathscr{P}_{GS1}$. If for some reason Pr chooses to counter B with F instead of G then, given $\mathscr{P}_{GS1}$, Pr wont be able to alter the outcome through backtracking, and thus will loose. In contrast, choosing to counter B with either D or F makes no difference for Pr under $\mathscr{P}_{GS2}$, since in either case backtracking allows for all possible choices to be exhaustively employed, and therefore if there exists a way for Pr to win, it will be reached.

Based on these observations Theorem 5 can be derived, to which we will be referring as the *Pinball Theorem*. Essentially, what the theorem states is that regardless of how a dialogue may evolve, if backtracking is allowed then the same outcome will always be reached. This is similar to how the balls of a pinball machine will always end falling into the drain situated at the bottom of the play field, regardless of the course they follow during the game.

For stating the theorem it is first necessary that we define a particular set of protocol rules that need to describe a dialogue for it to be characterised by the Pinball theorem.

Definition 40 (Protocol rules $\mathcal{P}_{\mathcal{S}}$) *Let $\mathcal{P}_{\mathcal{S}}$ stand for a set of protocol rules where:*

1. *The goals of the participants, Pr and Op, concern proving ($\mathcal{G}_{Pr}$) respectively disproving ($\mathcal{G}_{Op}$) the acceptability of an argument X, with respect to a given semantics $\mathcal{S}$={Credulous, Grounded};*

2. *The participants exchange arguments based on a binary attack relationship $\mathcal{C}$, Pr begins with X;*

3. *Backtracking is allowed, and so a dialogue $\mathscr{D}$ characterised by $\mathcal{P}_{\mathcal{S}}$ may be expressed in the form of a tree $\mathcal{T} = \{d_1, ...d_L\}$ with X as its root;*

4. *Assuming a labelling applied on $\mathcal{T}$ based on Definition 33, an argument can be introduced in $\mathscr{D}$ only if it changes the dialogical status of X;*

5. *If $\mathcal{S}$=Grounded then $\forall d \in \mathcal{T}$, no two arguments with speaker Pr are the same, while if $\mathcal{S}$=Credulous then $\forall d \in \mathcal{T}$, no two arguments with speaker Op are the same, and;*

6. *The game terminates if one of the players cannot respond in any way.*

Theorem 5 (Pinball Theorem) *Let there be a terminated persuasion dialogue $\mathscr{D}_1$ for an argument X characterised by $\mathcal{P}_{\mathcal{S}}$, and assuming that all arguments distinctly known to either participants that could have been moved in $\mathscr{D}_1$ have been moved, then if:*

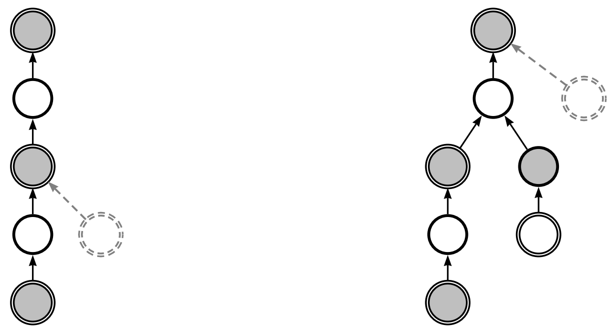
$$\texttt{result} : \mathscr{D}_1 \times \mathcal{G}_{Pr} \longrightarrow \{Success\}$$

then there exist no other sequence of dialogue moves $\mathscr{D}_2$ for X, subjected to the same assumptions that characterise $\mathscr{D}_2$ such that:

$$\texttt{result} : \mathscr{D}_2 \times \mathcal{G}_{Pr} \longrightarrow \{Failure\}$$

Respectively for $\mathcal{G}_{Op}$.

Proof We prove this by contradiction. For convenience let $I \in \{Pr, Op\}$ and bp_I represent a participant's branching—backtracking—point in a dialogue tree, i.e. a point where more than a single response choice are available to a participant.



(a) (b)

Figure 3.20: a) An example of Case A) where Pr (grey) has no branching choices, while Op (white) has an alternative which was not employed in the game, b) An example of Case B) where though Pr does have possible branching alternatives, they all follow after Op's branching point at the top.

Let us assume that a persuasion dialogue $\mathscr{D}_1$ for an argument X, terminates with $Pr(Op)$ as the winner. Let us also assume that there exists another sequence of moves $\mathscr{D}_2$ for X with the opposite result. Termination of $\mathscr{D}_2$, implies that a point is reached where:

1. $Pr(Op)$ cannot counter Op's(Pr's) last move, and;

2. $Pr(Op)$ cannot backtrack to an alternative reply.

Given that all disputes in a dialogue tree share *the same root*, as it is the case for the two concerned dialogues, it is then implied that:

A) either, $Pr(Op)$ had no branching options in $\mathscr{D}_1$ at all (advise the example in Figure 3.20a), or;

B) $Pr(Op)$ had possible branching points in $\mathscr{D}_1$ (advise the example in Figure 3.20b).

Let A) be the case, then:

i) $Op(Pr)$ must have had at least one branching point in $\mathscr{D}_1$, which, though not used, supports the existence of $\mathscr{D}_2$

ii) Existence of $\mathscr{D}_2$ suggests that the result of $\mathscr{D}_1$ was a 'false' choice—selection of argument—at a certain branching point by $Op(Pr)$, which led to $Pr(Op)$ being the winner.

iii) In turn, ii) implies that $\mathscr{D}_1$ should not have terminated when it did, since $Op(Pr)$ had a backtracking choice to a different branch, which contradicts with the assumption that $\mathscr{D}_1$ was terminated.

Let B) be the case, then:

i) if there were any branching options for $Pr(Op)$ in $\mathscr{D}_1$, then $Op(Pr)$ must have had at least one $bp_{Op}(bp_{Pr})$ in $\mathscr{D}_1$ preceding those of Pr's(Op's), through which $Op(Pr)$ would be able to guarantee its victory in $\mathscr{D}_2$.

ii) Note that in the opposite case, since we are concerned with a possible victory by Op, it would not be possible to guarantee that $\mathscr{D}_2$ does not contain a winning strategy for Pr, since:

   1) if $Op(Pr)$ had no branching options preceding those of Pr, then $Pr(Op)$ would primarily be the one able to affect the course of the dialogue game, diverting it back to the course followed in $\mathscr{D}_1$.

   2) 1) implies that it would then be impossible for $Op(Pr)$ to deviate from D_1, as there would be no move capable of altering the labelling status of the root argument, and thus to invalidate the existence of a winning strategy in $\mathscr{D}_1$.

   3) Given 2) and since there exists a terminated game that led to $Pr's(Op's)$ victory ($\mathscr{D}_1$) then, given backtracking, $Pr(Op)$ should always be able to reach the same outcome.

   4) However, assuming termination of $\mathscr{D}_2$, $Pr(Op)$ should not be able to reach the outcome achieved in $\mathscr{D}_1$ through backtracking.

iii) This implies that $\mathscr{D}_2$ resulted from an Op's(Pr's) branching point that precedes all of Pr's(Op's) branching points.

iv) Therefore the result of $\mathscr{D}_1$ must again be a 'false' selection of argument at a certain bp_{Op} in $\mathscr{D}_1$ by $Op(Pr)$, leading to $Pr(Op)$ falsely being the winner.

v) This suggests that $\mathscr{D}_1$ should not have terminated when it did since $Op(Pr)$ had a backtracking choice to a different branch, which was not employed.

vi) The latter contradicts with the assumption that $\mathscr{D}_1$ was terminated.

In general we have shown that the existence of $\mathscr{D}_2$ contradicts with the fact that its outcome should have already been reachable through $\mathscr{D}_1$, since both dialogues share the same root. Given backtracking, termination guarantees that all possible alternatives have been exhaustively countered.

3.4.2 Accounting for the Underlying Logic

As shown in Section 3.3.2, abstract approaches fail to account for the possibility of a new argument being instantiated from the logical content accumulated from the dialogue process. The instantiation of such an argument is evident in the example presented in Figure 3.5. However, apart from accounting for logical completeness when strategising, in the following example we illustrate another way of accounting for the structured form of arguments when which relates to the use of an OM, also not captured by abstract approaches.

In similar sense to the previous example, we assume a pair of agents engaging in the previous section's grounded game dialogues, making use of their beliefs about their interlocutor's knowledge to strategise, thus instantiating the general definition of the strategy function in Definition 27. Given a dialogue $\mathscr{D} = < \mathcal{DM}_0, \ldots, \mathcal{DM}_k >$ the protocol rules determine a set of possible locutions for inclusion in DM_{k+1}. `Speaker`$(\mathcal{DM}_{k+1})$ then relies the protocol; its own argumentation theory; the commitment store of `Hearer`$(\mathcal{DM}_{k+1})$, and; what it believes is `Hearer`$(\mathcal{DM}_{k+1})$'s argumentation theory; to simulate how the dialogue may evolve with respect to the possible locutions $\mathcal{M}_1, \ldots, \mathcal{M}_n$ for inclusion in DM_{k+1}. For $i = 1, \ldots, n$, the agent `Speaker`$(\mathcal{DM}_{k+1})$ simulates a dialogue $\mathscr{D}_i$ extending $\mathscr{D}$ with $\mathcal{DM}_{k+1}i$ containing locution $\mathcal{M}_i$.

The agent can then evaluate which of these dialogues result in success, and so make the choice of locution accordingly. Let us illustrate with an example.

Example 2 *Suppose an argumentation system* $(\mathcal{L}, ^-, \mathcal{R}, \leq)$ *where:*

- $\mathcal{L}$ *is a language of propositional literals, composed from a set of propositional atoms* $\{a, b, c, \ldots\}$ *and the symbols* $\neg$ *and* $\sim$ *respectively denoting strong and weak negation (i.e., negation as failure).* α *is a strong literal if* α *is a propositional atom or of the form* $\neg\beta$ *where* β *is a propositional atom.* α *is a wff of* $\mathcal{L}$ *if* α *is a strong literal or of the form* $\sim \beta$ *where* β *is a strong literal.*

- *For a wff* α*,* α *and* $\neg\alpha$ *are contradictories and* α *is a contrary of* $\sim \alpha$*.*

Consider a grounded persuasion dialogue $\mathscr{D} =< \mathcal{DM}_0, \ldots, \mathcal{DM}_k >$ *is to be extended with a Pr move* $\mathcal{DM}_{k+1}$*, and let us assume that for Pr to win, the dialectical labelling of* $\mathcal{DM}_k$ *must be made* `out`*. Figure 3.21a (resp. 3.21b) shows Pr's own (resp. what Pr believes is Op's) premises, rules, pre-orderings and goals, relevant to extending* $\mathscr{D}$*. Accordingly, Figures 3.21c and 3.21d, illustrate the set of arguments we assume that Pr can construct based on* $S'_{(Pr,Pr)}$*, and Pr's beliefs about the arguments that Op can construct based on* $S'_{(Pr,Op)}$*. Notice that because of the absence of premise* $p \notin \mathcal{K}_{(Pr,Op)}$*, Pr believes that Op is unable to instantiate argument E (Figure 3.21e). Pr has a choice of replying to* $\mathcal{DM}_k$ *with arguments A or* A' *and thus simulates the following two dialogue trees* $(\mathcal{T}1, \mathcal{T}2)$ *based on them, of which the second is depicted in Figure 3.21f* [1]*:*

• $\mathcal{T}1$*: Pr moves A leading to an immediate victory for Pr since it cannot be countered*

• $\mathcal{T}2$*: Pr moves* A'*. Op then replies with B, giving Pr a choice between C and D. In its simulation Pr opts for D, which leads to a repetition of B by Op as licensed by protocol* $\mathscr{P}_G$*, followed by Pr replying with C. Pr wins this dispute, making* $\mathcal{DM}_k$ `out` *(this would also have been the case if Pr had chosen C rather than D at the earlier choice point). Since p is now in Pr's commitment store, then given Pr's beliefs about Op's knowledge, Pr simulates Op's use of this commitment to construct E and simulates Op's backtrack moving E in reply to* A'*, thus making* $\mathcal{DM}_k$ `in`*. Hence Pr backtracks to move A, followed by Op reusing E which results in* $\mathcal{DM}_k$ `in`*. Based on* $\mathscr{P}_G$*, Pr cannot repeat A against E and thus loses the game*

[1] `in` and `out` labelled nodes are expressed with double respectively single lines. Dashed arrows concern the possible replies that may follow after a dialogue move

$S'_{(Pr,Pr)}$	$\mathcal{K}$	s, p, r, w
	$\leq'$	$\varnothing$
	$\mathcal{R}$	$s \Rightarrow a, r \Rightarrow a, p \Rightarrow q$
	$\leq$	$s > w$
	$\mathcal{G}$	$prove \;\; \alpha$

(a)

$S'_{(Pr,Op)}$	$\mathcal{K}$	w
	$\leq'$	$\varnothing$
	$\mathcal{R}$	$w, p \Rightarrow \neg a, \sim q \Rightarrow \neg r$
	$\leq$	$w > s$
	$\mathcal{G}$	$disprove \;\; a$

(b)

$\mathcal{A}'_{Pr}$	$A : s; s \Rightarrow a$
	$A' : r; r \Rightarrow a$
	$C : p; p \Rightarrow q$
	$D : r$
	$G : s$
	$F : p$

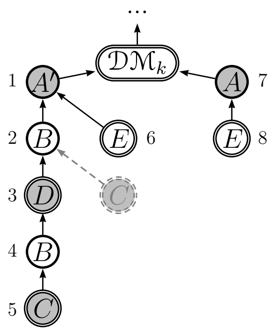
(c)

$\mathcal{A}'_{Op}$	$B : \sim q \Rightarrow \neg r$

(d)

$E : w, p; w, p \Rightarrow \neg a$

(e)



(f)

Figure 3.21: a) & b) illustrate a subset of Pr's own knowledge ($S'_{(Pr,Pr)} \subseteq S_{(Pr,Pr)}$) respectively beliefs about Op's knowledge ($S'_{(Pr,Op)} \subseteq S_{(Pr,Op)}$), c) & d) concern respectively the set of arguments $\mathcal{A}'_{Pr} \subseteq \mathcal{A}_{Pr}$ that Pr can construct based on $S'_{(Pr,Pr)}$, and the set of arguments $\mathcal{A}'_{Op} \subseteq \mathcal{A}_{Op}$ that Pr assumes Op can construct based on $S'_{(Pr,Op)}$, e) Argument E f) the simulated dialogue tree ($\mathcal{T}2$) instantiated if $DM_{k+1} =< Pr, A' >$

The above example illustrates that if Pr, in its strategising, accounts for the logical content of arguments updating the commitment store, the choice of content for $\mathcal{DM}_{k+1}$ makes a difference to the outcome of the actual dialogue, under the assumption that Pr's beliefs about Op's knowledge is indeed accurate. Pr prefers to move A rather than A', as the latter would result in there being no winning-strategy for Pr.

If one relied on an abstract representation of the employed arguments, the simulated dialogues (we show only the arguments) would have been $< A >$, and $< A', B, C >$ or $< A', B, D, B, C >$, all of which would make $\mathcal{DM}_k$ `out` and Pr winning. In other words, Pr would be indifferent to choosing between A and A' since the construction and use of argument E would not have been simulated. Given this we also argue that although soundness and fairness can be shown for the purely abstract approach, such an approach is inadequate, as it fails to accommodate the fact that new arguments can be made available during the course of a dialogue, due the dynamic evolution of knowledge available for argument construction (as shown by the use of the commitments of one agent in

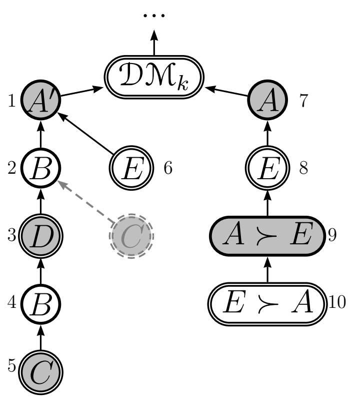


Figure 3.22: Example 2 extended with the use of preferences

the arguments constructed by another).

Further in relation to the provided example and to the use of preferences, assume that after the deployment of E by Op against A, Pr updates its preference-orderings such that not only $s > w$, but also $p \Rightarrow a > w, p \Rightarrow \neg a$. The latter would, under the weakest link principle, give the argument ordering $A \succ E$, which Pr can simulate moving in as a reply to E, thus making $\mathcal{DM}_k$ `out`. In a similar sense, suppose Pr assumes that Op will also update its preference-orderings such that in addition to $w > s$, Op also believes $p > s$ and $w, p \Rightarrow \neg a > p \Rightarrow a$. Then, again under the weakest link principle, this would produce a counter argument ordering $E \succ A$, which if included in Pr's simulation as Op's reply to $A \succ E$, and based on $\mathscr{P}_G$ which dictates that Pr cannot repeat $A \succ E$ in the same dispute, will result in the dialogue ending with $\mathcal{DM}_k$ `in`, as illustrated in Figure 3.22.

3.4.3 Other Strategic Considerations

Additional strategic considerations may concern external factors that could possibly affect a participant's goals, such as time dependencies. In the presence of time constraints a participant may require to reach a result as soon as possible and thus to probably take a risk. Similarly, in their absence, stalling the dialogue game for as long as possible may also be an objective since such an attitude could possibly result in protruding their interlocutor into taking a risk. Additionally, a participant's sole objective might concern exploring its interlocutors beliefs. In this sense stalling a dialogue game as long as possible may also serve in increasing the possibility of coming across new information about one's opponent.

These objectives are concerned with the structure of a dialogue tree. Particularly, they concern the depth of a winning line of dispute (tree path), the differentiation between winning and losing sub-trees and accordingly a calculation of a risk factor with respect to possible time dependencies or other similar constrains. Therefore, a strategic analysis of the structure of a simulated dialogue tree is imperative, in order to reveal any possible strategies that may be employed. Such issues could be accounted by a utility evaluation function and are further discussed in Chapter 6.

3.5 Conclusions & Contributions

In this chapter we have provided a general framework for dialogue that enables formal, off-line, analysis that each agent may undertake in order to strategise over the choice of moves to make in a dialogue game. The latter relies on a logical conception of arguments, that an agent may undertake in order to strategise over the choice of moves to make in a dialogue game, based on its model of its opponents.

Our intention was to make two main contributions in this chapter. Firstly, we defined persuasion dialogue instances of our framework, related to those described in Prakken [2005], but extended so as to also account for admissible semantics, while we provided soundness and fairness results for both. Furthermore, we have enabled the introduction of preferences into the game that 'undermine' the success of attacks as defeats relying on a similar formalism to the one proposed by Amgoud and Cayrol [2002b], though in the context of dialogues. These preferences may be contradictory and are effectively treated as mutually attacking arguments. The latter we consider to be a novel property of our system, while it suggests future work, building on Modgil [2009], to enable agents to argue about their preferences. We note other mechanisms have been proposed in the literature for describing and embodying preferences. The work of Amgoud and Cayrol [2002b] is only one such approach. Of some relevance, is the work of Bench-Capon et al. [2007]; Dunne and Bench-Capon [2004] where a similar formalism is proposed, though in the context of value-based argumentation (Bench-Capon [2003]). In their work "preference" is considered in terms of the notion of "audience": the relative importance that agents may ascribe to the (qualitative) social/ethical values associated with arguments, while agents are allowed to argue over value orderings on arguments for resolving conflicts.

The second concerns the provision of an $ASPIC^+$-based dialogue framework which allows for a structural analysis with respect to the underlying logic. We contrasted such an analysis with abstract opponent modelling, showing that appropriate mechanisms for strategising need to account for the logical content of arguments, thus accounting for the possibility of a new argument possibly being constructed from the mutually provided information, i.e. account for the logical completeness property. In addition, and particularly in relation to strategising we also showed how revealing certain constituents of an argument may allow one to instantiate arguments by combining its own as well as information provided from the game which could alter an anticipated result. We thus illustrated how the inherent feature of the dynamic construction of arguments can be captured though our framework, both with respect to the soundness of the framework as well as from a strategic perspective.

Overall, our framework not only accounts for this inherit nature of dialogues, but we also provide the means through which one may actually instantiate possible new arguments and include them in the dialogue process. Finally we note that because $ASPIC^+$ explicitly models the logical content and structure of arguments, while accommodating many existing logical approaches to argumentation, we can claim a similar level of generality for our dialogical framework.

A limitation of our framework is that it does not account for how agents update their own knowledge after the end of a dialogue game. This issue could be investigated under two perspectives. One concerns whether the loser of a dialogue should incorporate the disputed argument (we refer to its logical constituents) into his knowledge provided that he has failed to disprove its acceptability. Consequently one would find reasonable to also include arguments used by one's opponent if they are found within a winning-strategy in a dialogue tree. The other perspective could be concerned with whether the winner of a game should or should not incorporate any of his knowledge with arguments moved by his opponent.

One could support that all knowledge should be incorporated in both of the participants' actual knowledges, since it should not alter their personal perception of that argument's acceptability, provided that their personal perception coincides with the result of the dialogue game. However, this requires that agents are 'truthful' in their efforts to prove or disprove the acceptability of a disputed argument. In other words, that they introduced into the game all possible arguments they could have introduced, and did not conceal any information due to pursuing side objectives. In addition, the introduction of a set of arguments may affect the acceptability status of other arguments in an agent's own knowledge-base and this may be an undesirable side-effect. A more flexible approach would be to allow the participants to either *concede* to their defeat or *reject* it and respectively choose to update or not update their knowledge.

These are interesting issues that we intend to investigate in future work since they are out of the research scope of this thesis. We simply note that, for the purpose of this thesis, regardless of whether agents do or do not update their knowledge with knowledge provided by their opponents or acquired in any other way (such ways are listed in the next chapter), we assume that every agent's set of arguments—arguments instantiated from their own sub-theories ($S_{(i,i)}$)—increase monotonically. This is compatible with the idea that the beliefs from which arguments are constructed are not revised upon incorporation of conflicting beliefs, but rather the conflicts are resolved through evaluation of the justified arguments under acceptability semantics (Dung [1995]).

Chapter 4

Opponent Modelling

For a participant in a dialogue, the *best response problem* concerns the selection of a locution that (by some measure) *optimises* the outcome. As Dunne and McBurney [2004] explain, the precise interpretation of 'optimise' may vary greatly, depending on the nature and intent of a dialogue area. Within a competitive context of persuasion dialogues a perspective of this problem concerns the strategic choice of a locution to make, from amongst a range of possible locutions, so as to better achieve its objectives. Many researchers rely on a participant's modelling of their opponents, when modelling the participant's strategy. In the literature this is widely studied in terms of *opponent modelling* in competitive game contexts. In most cases though, the formalisation of an *opponent model* is left implicit. The basic assumption is that knowledge about one's opponents is obtained as a result of accumulated experience in playing against them, or may be supplied by some external source. However, if one is to rely on such a model for strategising, it is necessary to formally define the mechanisms responsible for building and updating it, so as to account for a variety of factors that may have a considerable impact on the model's credibility, which in turn largely affects the effectiveness of any strategic decision. In this respect, an interesting research direction is to investigate how an opponent model may be built and updated so as to achieve a high level of credibility.

In this chapter we assume the general framework presented in Chapter 3 focusing on persuasion dialogues, and rely on a structural conception of arguments to formally define a number of mechanisms responsible for building, updating, and augmenting an opponent model. We focus on the augmentation mechanism, which is used to expand an agent's current model of its opponent, by adding to it information that has a high likelihood of being related with information already contained in it. Precise computation of this likelihood is exponential in the volume of related information. We thus describe and evaluate an approximate approach for computing these likelihoods based on *Monte-Carlo* simulation.

4.1 Introduction

The essential contribution of the dialogue framework presented in Chapter 3, is not only that it allows for the dynamic nature of dialogues to be taken into account, thus guaranteeing the system's soundness and fairness with respect to the underlying logic, but that it also provides the means based on which this essential property of dialogues can be taken into account when strategising. As we have extensively discussed, not accounting for the dynamic instantiation of an argument through the dialogue process, results in possibly providing one's interlocutor with crucial information based on which a new argument could be instantiated and introduced into the game, possibly altering its outcome. However, even accounting for this dynamic property of dialogues does not change the fact that the effectiveness of any strategy, largely depends on the credibility of one's opponent model (OM).

We begin this chapter in Section 4.2, by presenting a simple mechanism for building and updating an OM relying on an agent's experience in playing against a certain adversary. Since such problems are inherently part of a wider class of decision problems with which game theory is concerned, we borrow certain game theory concepts for providing a theoretical basis for our research, beginning with defining the notion of *history*, which is generally concerned with the logical information exchanged between players in dialogues.

In Section 4.2.2 we define and associate a *confidence* value with the logical information found in an OM, based on how that information became part of the model, i.e. based on the *information collection methods (ICMs)* which may be employed by an agent. We then explain how this value is used for characterising the credibility of an OM and assists in accounting for the fact that an agent may be in error in its modelling, as well as for resolving possible conflicts which may occur in the modelling process. We differentiate between three distinct ICMs: direct collection through the opponent's commitment store in dialogues (on which the basic update mechanism relies on); information provided by a third party, and; information added as a result of an *augmentation* attempt of an OM.

We focus on the augmentation mechanism in Section 4.3, which essentially concerns how one may extend its beliefs (arguments) about its opponent's beliefs, based on what one already believes its opponent believes. The proposed technique relies on how arguments used in dialogues against a certain opponent are *likely* to be associated in some way with other arguments which appear in dialogues with other adversaries. This association is represented in the form of a directed graph referred to as *relationships graph ($\mathcal{R}G$)* (a graph of arguments), assumed to be incrementally constructed from an agent's general history of dialogues with all of its opponents, and not just on its experience against playing a certain opponent alone.

We discuss the complexity of this augmentation process, and propose a Monte-Carlo simulation in Section 4.4, which relies on experimentally computing the distinct likelihoods of the relations between arguments. We provide experimental results for the proposed methodology, prove convergence of the produced likelihood values towards the corresponding theoretical values while we also show how it significantly improves the tractability of our approach.

In Section 4.5 we show how based on the results of the Monte-Carlo simulations one can incorporate the constituents of the additional arguments into an OM, focussing on the resolution of confidence conflicts between the same constituents choosing between those already in the model and newly incorporated ones, and show how a confidence value can then be generated to characterise arguments instantiated from the OM. We finally summarise our work and our contribution and present our conclusions in Section 4.6.

4.2 The Modelling Framework

The notion of game theoretic history was first introduced by Osborne and Rubinstein [1994] and is concerned with a particular category of games called *extensive* *games*. Informally, an extensive game concerns a specification of the sequential structure of decision problems encountered by participants (players) in a strategic situation. In turn, a *sequential game* (or dynamic game) is one where the participants have some information about earlier actions of other participants. This information might either be *perfect*, meaning that all participants know the actions made by all other participants, or *imperfect* meaning that a certain subset of the previous moves of other participants is known to the players.

For example, the persuasion dialogue instance of the general framework presented in Section 3.2 can be described as an extensive game with perfect information, given its sequential—turntaking—structure and the employment of commitment stores. To recap, in Section 3.1, we assume each agent Ag_i operating in a multi-agent environment to have a model of its opponents, expressed as an agent theory $AgT_i =< S_{(i,1)}, \ldots, S_{(i,\nu)} >$. An agent theory consists of a number of sub-theories, each of which expresses an OM, $S_{(i,j)} = \langle AT_{(i,j)}, \mathcal{G}_{(i,j)} \rangle$ where $i \neq j$, concerned with a particular opponent agent (Ag_j) in a multi-agent environment. Each sub-theory expresses Ag_i's beliefs about Ag_j's argumentation theory $AT_{(i,j)}$, i.e. its premises ($\mathcal{K}$), pre-ordering on those premises ($\leq'$), rules ($\mathcal{R}$), pre-ordering on those rules ($\leq$), and goals ($\mathcal{G}$).

As discussed in the previous chapter, by modelling information based on a logical conception of arguments, we allow for a more thorough strategic analysis with respect to the underlying logic and its effects on the dialogue process. However, in contrast with the opponent modelling approaches presented by Carmel and Markovitch [1996] and Oren and Norman [2010] we do not accommodate the notion of nested OMs. In other words, we do not account for the possibility that one's opponent might also adapt its strategising based on its own OM of its interlocutors, i.e. to *counter-strategise*. We, nevertheless discuss this issue in Chapter 6, where we present a number of strategising approaches proposed in relation to the matter and show how one may extend our framework in order to account for *counter-strategising*.

4.2.1 The notion of History

In the context of dialogue games, and for an agent operating in a multi-agent environment, the notion of history should ideally include all information exchanged between an agent and its opponents in a series of distinct dialogues. In addition, it should also include information related to when, or against which other moves, was a particular move deployed in a game, i.e. account for the structure of dialogues, as the latter encodes information related to a participant's strategy. We thus define a history as follows:

Definition 41 (History) *Let Ags be a set of agents* $\{Ag_1, \ldots, Ag_\nu\}$*, then* $\forall$ Ag_i*,*Ag_j*,* $j \neq i$*, we assume* $h_{(i,j)} = \{\mathscr{D}^1, \ldots, \mathscr{D}^k\}$ *to be* Ag_i*'s history of dialogues with* Ag_j*. Then:*

$$\mathcal{H}_i = \bigcup_{j=1, j\neq i}^{\nu} h_{(i,j)}$$

is the set of all histories of Ag_i *with each member* $Ag_j \in Ags$*.*

One of the aspects this chapter is concerned with how an agent's experience can be exploited in a multifaceted way. In other words we care to learn more from an agent's history of dialogues than just the locutions deployed by its opponents. We thus focus on the order based on which these locutions are exchanged. As the latter is captured by the tree-structure of dialogues, and is thus part of an agent's history $\mathcal{H}$, we propose a function below which is specifically concerned with the *distinct* disputes found in a set of dialogues between two agents, which, as it will become more evident later, better fits our modelling objectives.

Definition 42 (Disputes' History) *Let* $h_{(i,j)}$ *be* Ag_i*'s* ***history*** *of dialogues with* Ag_j*, then:*

$$\boldsymbol{\mathtt{disputes}}(h_{(i,j)}) \to \{d_m | \exists \mathscr{D}^k \in h_{(i,j)} : d_m \in \mathscr{D}^k\}.$$

In other words, `disputes` is a function that returns the set of all distinct disputes found in all dialogues between agents Ag_i and Ag_j.

Having defined the notion of an agent history $h_{(i,j)}$ one can better understand the nature of an OM ($S_{(i,j)}$) which basically encapsulates all the logical elements found in the dialogues of a history, divided in distinct subsets based on their nature. The difference between an OM and an agent history is that tan OM ignores the structure of the dialogue from which information is collected. In other words, it ignores the *where*, the *when* and *why* were certain logical elements introduced into a game through an argument. Thus, any strategy that may be developed on the basis of an OM consequently lacks in accounting for these questions as well. One of our research objectives is to propose a modelling approach able to additionally account for the structure of dialogues from which logical information is collected.

Given an agent's history one can instantiate and maintain an OM through the employment of an update mechanism whose form depends on the nature of the information collection method (ICM) for which it is employed. We proceed with providing a set of definitions of update mechanisms. These are particularly concerned with the different ICMs we care to investigate for the OM defined in Definition 17 in Chapter 3, which focusses on the logical information exchanged in dialogues, and and disregards modelling an agent's opponents' strategies.

At this point it is worth discussing that modelling an agent's opponents' strategies, e.g. by deducing possible behavioural patterns of moves in a series of dialogues, may not always be possible, or, even if it is, it might not be as fruitful as one would expect. This is because agents participating in dialogues are assumed to rely on a dynamic way of strategising for satisfying their current objectives in a game. Note that apart from persuading its opponent in a dialogue, a participant may also have the objective of stalling, or gaining information about, deceiving or even misleading its opponent. It might even be that agents alter their objectives half way through a dialogue game. In essence, in every dialogue a participant is only concerned with achieving the highest possible utility with respect to its goals, and thus accordingly adjusts its strategy in the game. Therefore, as Oren and Norman [2010] argue, for modelling an agent's current strategy it is enough that we deduce its current goals. In this sense, monitoring an agent's behaviour in a series of games could assist in deducing its goals in a current game. However, given that the we do not provide an explicit set of the possible goals that an agent might pursue in a dialogue game, we leave the goals update function implicit, and also leave this particular topic for future investigation.

4.2.2 Information Collection Methods

We begin by associating a *confidence* value c to the elements of the sets:

$$< \mathcal{K}, \mathcal{R}, \leq', \leq, \mathcal{G} > \quad of \quad S_{(i,j)}.$$

Essentially, for an agent Ag_i this value expresses the probability of a certain belief being part of Ag_j's actual knowledge ($S_{(j,j)}$). To compute this value we differentiate between whether a particular information is:

a) gathered directly by Ag_i, on the basis of its opponents' updated commitment store, referred to as `dir`,

b) provided by a third party, referred to as `tp`, or

c) a result of an *augmentation* attempt of Ag_i's current model of Ag_j, referred to as `aug`.

The latter (`aug`) concerns an incrementation of a current OM with the addition of beliefs that are *likely* to also be known to Ag_i's opponent, which is further analysed in this Section 4.3.

In the case of `dir`, we will generally assume that every agent always *retains* its own beliefs without revision, i.e. that introduction of conflicting information *does not cause any older information to be discarded*, but rather conflicts are decided by some non-monotonic inference mechanism—in particular the non-monotonic inference relation defined by the $ASPIC^+$ argumentation theory, as the latter is described in Chapter 3. We therefore assume that the confidence value c of information acquired directly from the commitment store of one's opponent is equal to 1, which represents the highest level of confidence.

Still, avoiding revision of the modelled knowledge base can hold only for all the sets of logical components comprised in the distinct sets of an agent's knowledge ($\mathcal{K}, \leq', \mathcal{R}, \leq$), but not for goals ($\mathcal{G}$), as it is unreasonable for an agent to simultaneously pursue conflicting objectives. However, as discussed, we leave the goals' part implicit.

4.2.2.1 Direct Collection

We proceed with defining four distinct update mechanisms with respect to each of the logical components $\mathcal{K}, \leq', \mathcal{R}, \leq$, found in a sub-theory S of an agent's AgT. To do so we rely on the notion of a commitment store as the latter is generally described in Definition 23, and provide a more explicit definition for each of the sets in S.

Definition 43 (Commitment stores) *Given a set of agents $\{Ag_1, \ldots, Ag_\nu\}$ participating in a dialogue $\mathscr{D}$, and assuming that σ_i, where $i = 1, \ldots, \nu$, is a subsequence of $\mathscr{D}$ which lists all moves by agent Ag_i in the order they appeared in $\mathscr{D}$, such that $\sigma_i =< \mathcal{DM}_0, \ldots, \mathcal{DM}_n >$, then for any agent Ag_i a* commitment store *is a tuple:*

$$CS_i =< \mathcal{K}_i, \leq'_i, \mathcal{R}_i, \leq_i >$$

whose evolution is defined as follows:

1. $CS_i^0 = < \varnothing, \varnothing, \varnothing, \varnothing >$

2. For $k = 0 \ldots n, CS_i^{k+1}$ is obtained by updating CS_i^k with the contents of the dialogue move $\mathcal{DM}_k$ based on the following definition of an update function $\mathscr{U}_{CS}$:

$$\mathscr{U}_{CS}(CS_i^k, \mathcal{DM}_k) \longrightarrow$$

$$\begin{cases} CS_i^{k+1} =< \mathcal{K}_i^k \cup \texttt{Prem}(X), \leq'^k_i, \mathcal{R}_i^k \cup \texttt{Rules}(X), \leq_i^k > & \text{if } \mathcal{DM}_k^{con} = X \in \mathcal{A} \quad (a) \\ CS_i^{k+1} =< \mathcal{K}_i^k, \leq'^k_i \cup X, \mathcal{R}_i^k, \leq_i^k > & \text{if } \mathcal{DM}_k^{con} = X \subseteq \leq'_i \quad (b) \\ CS_i^{k+1} =< \mathcal{K}_i^k, \leq'^k_i, \mathcal{R}_i^k, \leq_i^k \cup X > & \text{if } \mathcal{DM}_k^{con} = X \subseteq \leq_i \quad (c) \end{cases} \tag{4.1}$$

Provided these and upon completion of a dialogue game, an agent's OM is updated based on the following definition:

Definition 44 (Update mechanism) *For an agent Ag_i, $\mathscr{U}$ is an update function that takes as input the current version of an OM expressed as a sub-theory:*

$$S_{(i,j)}^{\mu-1} =< \mathcal{K}_{(i,j)}^{\mu-1}, \mathcal{R}_{(i,j)}^{\mu-1}, \leq'^{\mu-1}_{(i,j)}, \leq_{(i,j)}^{\mu-1}), \mathcal{G}_{(i,j)}^{\mu-1} >$$

and the commitment store:

$$CS_j =< \mathcal{K}_j, \leq'_j, \mathcal{R}_j, \leq_j >$$

of a dialogue $\mathscr{D}^{\mu}_{(i,j)}$ and produces a new sub-theory $S^{\mu}_{(i,j)}$ according to the following update functions:

$$\mathscr{U}(CS_j, S^{\mu-1}_{(i,j)}) \longrightarrow S^{\mu}_{(i,j)} = \begin{cases} \mathcal{K}^{\mu}_{(i,j)} = \mathcal{K}^{\mu-1}_{(i,j)} \cup \mathcal{K}j & (a) \\ \mathcal{R}^{\mu}_{(i,j)} = \mathcal{R}^{\mu-1}_{(i,j)} \cup \mathcal{R}j & (b) \\ \leq'^{\mu}_{(i,j)} = \leq'_j \cup \leq'^{\mu-1}_{(i,j)} & (c) \\ \leq^{\mu}_{(i,j)} = \leq_j \cup \leq^{\mu-1}_{(i,j)} & (d) \end{cases} \tag{4.2}$$

In essence, the four distinct update mechanisms presented here simply incorporate the newly obtained information into the existing logical sets.

4.2.2.2 Handling third party information

In relation to third party provided information, a modeller's existing OM will also be updated according to Definition 44. That is, premises, rules, as well as priority ordering over those premises and rules, will simply be added to the corresponding sets in the existing model. Nevertheless, such information cannot be characterised with the same level of confidence that characterises information directly collected by the modeller, as trust issues arise. We have to therefore develop a different way of handling third party information through the development of a methodology for assigning a confidence value to it.

Relying on third party information for strategising is not a trivial issue, as it requires the handling of trust issues and deception. Assume for example that an agent, George, informs another, Michael, that a third agent, Mark is aware of a certain argument. To what level should Michael trust that this information is valid? In other words issues concerned with the credibility of the provided information, the credibility of the source that provides that information, or even whether certain information is provided by a third party as an attempt to mislead or deceive its own opponents, need to be resolved.

In this respect, it is important that trust mechanisms are employed, so as to deal with this additional kind of uncertainty. Intuitively, what needs to be handled is how the credibility of information changes as it is passed from one agent to another. A common approach for dealing with this problem is to associate that information with a credibility value, normalised to express the probability of that information to actually hold (Y. Tang et al. [2010]). Of course this results in assigning a similar credibility (confidence) value to all the information providers operating in a multi-agent environment—a degree of trust. Consequently, the credibility of a particular information provided by an agent Ag_i, to which it was provided by another agent Ag_j, to which it was in turn provided by an agent Ag_z and so on, can be computed by means of calculating the propagated trust value of that information. The latter results from following a path beginning from the information provider and ending, through the intermediate providers, to the original source.

Assume for example that information is passed between four co-workers $Mark$, $John$, $Michael$ and $George$. John believes that Mark is aware of an argument A with a certain confidence value equal to c, and that this information is conveyed to $Michael$ who trusts $John$ with a certain trust value $tr_{(Michael,John)}$ who then passes it to $George$ who trusts $Michael$ with a $tr_{(George,Michael)}$. Assuming that both the confidence as well as the trust values are normalised values which essentially represent probabilities, then computing the trust value of the conveyed information from George's perspective, could be done through multiplying the trust values found on the path between John and and George, as well as with John's confidence value on argument A, e.g.:

$$tr^c_{(George,John)} = c \cdot tr_{(Michael,John)} \cdot tr_{(George,Michael)}$$

since these quantities are assumed to be dependent. The propagated probability value $tr^c_{(George,John)}$ can then be associated with the third party information as a confidence indicator for the conveyed information.

For doing this, one may rely on a *relationship network* (an example of the latter is provided in Y. Tang et al. [2010]). As discussed, for the purpose of our work we incorporate the idea of assigning a numerical credibility value to all parts of logical information concerned with an OM as well, expressed in the form of confidence. Thus in the case where information is provided by a third party, the confidence of that information will be equal to the propagated trust value of that information represented as $tr_{(Ag_i,Ag_j)}$ according to the following definition:

Definition 45 *Let* $T_N = \{Ags, \tau\}$ *be a trust network where* $Ags = \{Ag_1, \ldots, Ag_\nu\}$

is a set of agents in a multi-agent environment and τ is a set of binary trust relations over the agents in Ags such that if $(Ag_i, Ag_j) \in \tau$ then $\{Ag_i, Ag_j\}$ is a directed arc in T_N. Let $p =< Ag_i, Ag_{i+1}, \ldots, Ag_{j-1}, Ag_j >$ be a path between agents Ag_i and Ag_j in a T_N, while we assume that Ag_j holds information on another agent Ag_w with a certain confidence c which is conveyed through p to Ag_i then:

$$tr^c_{(Ag_i,Ag_j)} = c \cdot tr_{(Ag_i,Ag_{i+1})} \cdot tr_{(Ag_{i+1},Ag_{i+2})} \cdot \ldots \cdot tr_{(Ag_{j-1},Ag_j)}$$

where $tr^c \mapsto [0, 1]$.

We opt to leave the formal definition of the trust values tr implicit, as there are already many numerical systems that deal with propagating trust information, such as those presented by Katz and Golbeck [2006]; Richardson et al. [2003], and Wang and Singh [2006]. We also ignore issues related with how these trust values may be moderated based on reputation information or based on a subjective evaluation of the credibility of a provider, as these are not in the scope of our work. We refer the reader to Section A.2 of Appendix A, where, borrowing from Y. Tang et al. [2010], we show how a trust network can be defined in a multi-agent environment, while we also provide an abstract function for computing the propagated trust value which we assume to be assigned in turn as a confidence value to the logical information of an OM acquired from a third party agent.

4.2.2.3 The Confidence Value

So far we have defined two different kinds of information collection methods (ICMs). These are direct collection of information from commitment stores and provision of information by a third party agents. The third party ICM inherently has to deal with deception, which is a very difficult issue to address. As in the case of "hidden agendas" (Silverman [2005]) third party agents may provide misleading information to a modeller in an attempt to satisfy self-interested objectives. We address this possibility only to some extent by relying on trust. We know however, that handling trust does not completely resolve deception issues, and this is something we intend to investigate in future work.

In relation to the direct collection of information ICM, we iterate that the employment of this method bears on the assumption that information is never discarded. Specifically, we assume that even in the case where an agent is faced with conflicting information, this is resolved through the employment of some non-monotonic inference mechanism. We however assume monotonicity of argument construction. In other words, we assume that agents will always be able to construct arguments they previously uttered. To some extent, this is an unrealistic assumption, since argument construction can be threatened by new information (e.g. information that does not allow the use of negation as failure any more, or information that undercuts the use of a certain defeasible rule). In addition, we cannot exclude the possibility of an agent simply deleting information for whatever reason (e.g. to save resources). Though we are aware of these issues, we consider them to be out of the scope of this thesis and leave them to future work.

In the following sections we will be dealing with another information collection method which concerns an *augmentation* attempt on an OM. Prior doing so, we first define how a modeller assigns confidence values to the collected information, to be later used for decision making purposes.

In general, based on the ICM through which information is collected it will accordingly receive a corresponding confidence value. This is described by the following definition:

Definition 46 (Confidence assignation) *Let $S_{(i,j)} \in AgT_i$, and $Y \in \{\mathcal{K}_{(i,j)}, \leq'_{(i,j)}, \mathcal{R}_{(i,j)}, \leq_{(i,j)}, \mathcal{G}_{(i,j)}\}$, then X is a tuple $< x, c >$ such that:*

- *$x \in Y$, and;*
- $c = \begin{cases} 1 & \textit{if } x \textit{ is directly collected by } Ag_i & (a) \\ tr^c_{(Ag_i, Ag_z)} & \textit{if } x \textit{ is provided by a third party agent } Ag_z & (b) \\ Pr(x) & \textit{if } x \textit{ is part of an augmentation of } S_{(i,j)} & (c) \end{cases}$

where c represents the confidence level of x, and where $Pr(x)$ is the likelihood of x being known to Ag_j, determined partly by the current form of $S_{(i,j)}$.

In case (c) of Definition 46 the confidence value $Pr(x)$ is assumed to be obtained by the augmentation process analysed in the next section, which aims to increase the credibility of an OM.

4.3 The Augmentation Mechanism

The basic idea behind the employment of an OM is its use for simulating the possible choices of one's opponent in a dialogue game. In contrast with perfect information games though, anticipating one's opponent moves in dialogue games is much more difficult, given that one cannot be perfectly aware of one's opponents knowledge. Thus strategising against an opponent needs to additionally account for the possibility that the opponent is aware of more (or less) alternatives, possibly not part of one's OM. In other words, a participant might be in error in its modelling of its opponents' beliefs, or may hold beliefs about its opponents' knowledge with varying degrees of certainty, and this is something that should also be accounted for.

Unfortunately, it is impossible to model opponent information perfectly, as we cannot know everything that our adversaries might use against us in a dispute, and this is a basic characteristic of the nature of disputes. However, a possible way based on which we may increase the credibility level of an OM is through addressing the question of how we can add further reliable information to it—*augment* it—based on:

a) what we currently believe our opponent believes, and

b) what we assume others, with knowledge similar to our opponent's, believe.

This can be done with the development of an *augmentation mechanism*, through which one may add to an OM information that is *likely*—has a high probability—to be associated with information already contained in it. For doing this we rely on an agent's general history of dialogues monitoring the times that certain opponent arguments (OAs) follow after certain others. In this way we utilise an agent's experience in a multifaceted way:

a) through updating an OM based on information an agent collects directly from a particular opponent through their history of interactions as described in Section 4.2.2, and

b) through relying on it for associating the logical information in an OM with other information, external to the model, based on an agent's general history of dialogues.

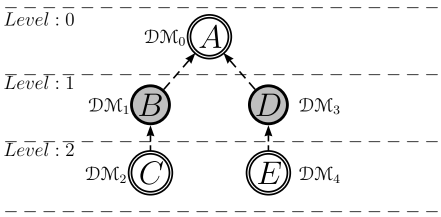


Figure 4.1: A dialogue between agents Ag_1 and Ag_2

The proposed approach is concerned only with persuasion dialogues defined in Section 3.2.

In order to convey the intuition behind our approach we begin with the following example:

Example 3 *Let two agents, Ag_1 and Ag_2, engage in a persuasion dialogue in order to decide where is the best place to have dinner:*

- *Ag_1:(A) We should go to the Massala Indian restaurant since a chef in today's newspaper recommended it.*

- *Ag_2:(B) A single chef's opinion is not trustworthy.*

- *Ag_1:(C) This one's is, as I have heard that he won the national best chef award this year.*

- *Ag_2:(D) Indian food is too oily and thus not healthy.*

- *Ag_1:(E) It's healthy, as it's made of natural foods and fats.*

The dialogue illustrated in Figure 4.1 is essentially composed of two lines of dispute, $\{A \leftarrow B \leftarrow C\}$ and $\{A \leftarrow D \leftarrow E\}$. Assume then, that Ag_2 engages in a persuasion dialogue with another agent, Ag_3, on which is the best restaurant in town. Let us also assume that at some point in the dialogue Ag_3 cites the newspaper article, by asserting argument A in the game, as Ag_1 did in the previous dialogue. It is then reasonable for Ag_2 to expect that to some extent Ag_3 is likely *to also be aware of the chef's qualifications (argument C).*

Intuitively, this expectation is based on a relationship between consecutive arguments in the same dispute lines of a dialogue. In this case, the "chef's proposition" (A) is defended against B's attack (i.e., 'supported') by "his qualifications" (C), suggesting some likelihood that awareness of the first implies awareness of the second. This is also the case for arguments A and E. However, assuming such a relationship between the chef's qualifications (C) and the argument on why Indian food is considered healthy (E), seems less intuitive, as these arguments belong in different dispute lines, and so do not support each other. We thus assume that two arguments can be related if they are found in the same dispute lines of a dialogue, where this relationship can be understood in terms of the notion of *support*, e.g., C supports A against B.

Given this, one may rely on an agent's accumulated dialogue experience to define a graph in which links between OAs asserted in a series of dialogues, indicate support. We will be referring to this graph as *relationships graph* ($\mathcal{R}G$). A participant may then rely on a $\mathcal{R}G$ to *augment* an OM, by adding to it arguments (precisely, their logical constituents) which, according to the graph, are linked with other arguments already contained in the OM (i.e. arguments which may be constructed from the logical elements found in an OM), and which are thus likely to be known to that opponent.

At this point it is worth going through a couple of detailed examples so as to illustrate how the proposed methodology can be found useful in the sense that it can increase the effectiveness of an agent's strategising. For presentation convenience we will avoid referring to the logical constituents contained in an OM from which one may instantiate a set of arguments and will assume that OMs simply contain sets of arguments. What we want to stress is the relationship of those arguments with other arguments found in a $\mathcal{R}G$, and how exactly that relationship might be of use to a modeller when found at a strategic point. In the following examples we investigate two scenarios where backtracking *is* and *is not* allowed, and where the sole objective is winning.

Example 4 (Backtracking not allowed) *Let $\mathscr{D}_k$ be an ongoing dialogue between two agents Ag_i (Grey) and Ag_j (White) where Ag_i has initiated a dialogue by moving argument A ($\mathcal{DM}_0$), found at the root of the dialogue tree presented in Figure 4.2a). Let us also assume that Ag_i holds an OM for Ag_j (Figure 4.2b)) which comprises arguments B and D thought to be part of Ag_j's knowledge base with absolute certainty. At the same time Ag_i is aware of arguments C, G and E.*

Let us now assume that the following scenario evolves:

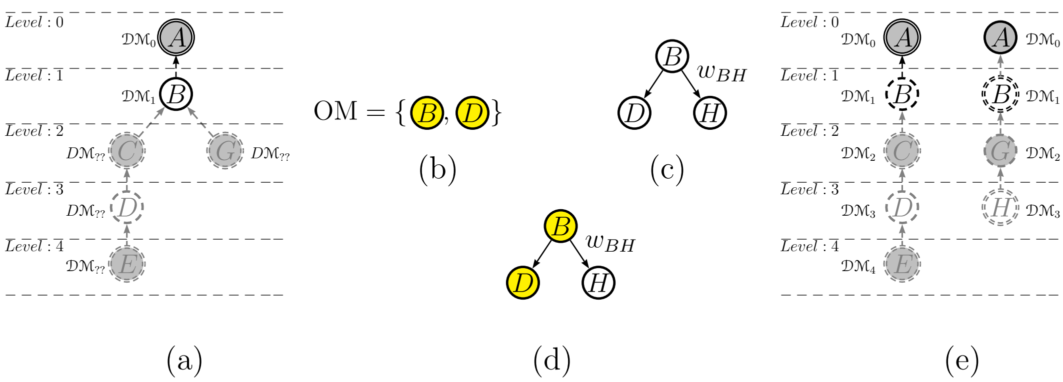


Figure 4.2: Example 4: a) A simulated dialogue based on just the OM, b) Ag_i's OM of Ag_j, c) Ag_i's $\mathcal{R}G$, d) The combination of Ag_i's OM of Ag_j's and its $\mathcal{R}G$, e) The extended simulated dialogue tree which results from the addition of OM's neighbouring arguments in the $\mathcal{R}G$.

1. *Let B be an attacker of A, and assuming that Ag_j does actually know B, then B is moved into the game as the content of its first move $\mathcal{DM}_1$ against $\mathcal{DM}_0$.*

2. *Assume then that both C and G can attack B, then Ag_i is found at a strategic point having to choose between the two arguments.*

3. *Given Ag_i's belief of D's existence in Ag_j's knowledge, and assuming that D is an attacker of C while additionally E is an attacker of D, Ag_i can simulate the two possible paths which describe how the dialogue may evolve. These are:*

   - $A \leftarrow B \leftarrow C \leftarrow D \leftarrow E$, *and;*
   - $A \leftarrow B \leftarrow G$.

   as they are both depicted in Figure 4.2a).

At this point the modeller, Ag_i, is relying solely on the OM in order to simulate the possible dialogue tree, according to which either option—C or G—is equally preferable, given that the objective of the game is simply winning and since both options lead to a winning leaf.

4. *Assume now that in addition to the OM (Figure 4.2b) Ag_i holds a $\mathcal{RG}$ (Figure 4.2c)), according to which arguments D and H usually follow after B in dialogues, with some likelihoods w_{BD} and w_{BH}, where w is a weight value expressed in the form of a probability.*

5. *As B is part of Ag_j's—the opponent's—knowledge, then based on Ag_i's $\mathcal{RG}$ it is likely that Ag_j is aware of H with some probability w_{BH} (Figure 4.2d)).*

6. *Given this, Ag_i can now extend the previous dialogue tree (Figure 4.2a)) through adding argument H following after G (Figure 4.2e[1]).*

Notice that, regardless of the probability w_{BH} which describes how likely it is for Ag_j to be aware of H, moving G into the game is no longer the clear option as it may *lead to a defeat, given that backtracking is not allowed.*

7. *Thus, Ag_i chooses argument C and moves it into the game.*

Example 5 (Backtracking allowed) *Let $\mathscr{D}_k$ be an ongoing dialogue between two agents Ag_i (Grey) and Ag_j (White) where Ag_i has initiated a dialogue by moving argument A ($\mathcal{DM}_0$), found at the root of the dialogue tree presented in Figure 4.3a. Let us also assume that Ag_i holds an OM for Ag_j (Figure 4.3b) which comprises arguments B and D thought to be part of Ag_j's knowledge base with absolute certainty. At the same time Ag_i is aware of arguments C, G and E.*

Let us now assume that the following scenario evolves:

1. *Let B be an attacker of A, and assuming that Ag_j does actually know B, then B is moved into the game as the content of its first move $\mathcal{DM}_1$ against $\mathcal{DM}_0$.*

2. *Assume then that both C and G can attack B, then Ag_i is found at a strategic point having to choose between the two arguments.*

3. *Given Ag_i's certainty of D's existence in Ag_j's knowledge, and assuming that D is an attacker of C while additionally E is an attacker of D, Ag_i can simulate the two possible paths which describe how the dialogue may evolve. These are:*

[1] The dialogue tree in Figure 4.2e is deliberately separated into two parts, since the resulting labelling is different for every case.

- $A \leftarrow B \leftarrow C \leftarrow D \leftarrow E$, *and;*
- $A \leftarrow B \leftarrow G$.

as they are both depicted in Figure 4.3a.

At this point the modeller, Ag_i, is relying solely on the OM in order to simulate the possible dialogue tree, according to which either option—C or G—is equally preferable, since both options lead to a winning leaf.

4. *Assume now that in addition to the OM (Figure 4.3b)) Ag_i holds a $\mathfrak{R}G$ (Figure 4.3c), according to which arguments D and H usually follow after B and F follows after D, with likelihoods equal to w_{BD}, w_{DF} and w_{BH} respectively, where w is a weight value expressed in the form of a probability.*
5. *As B is part of Ag_j's—the opponent's—knowledge, then based on Ag_i's $\mathfrak{R}G$ it is likely that Ag_j is aware of H with some probability w_{BH} (Figure 4.3d).*
6. *Furthermore, given D's relationship with F in the $\mathfrak{R}G$ and since Ag_j is assumed to be aware of D according to the OM, it is also likely that Ag_j is aware of F with some probability w_{DF} (Figure 4.3d).*
7. *Given this, Ag_i can now extend the previous dialogue tree (Figure 4.3a) through adding arguments H and F following after G and D respectively (Figure 4.3e).*

Notice that whereas before either choice was leading to victory now both options seem to be leading to defeat. Thus, once again the modeller is found at a point where he may be indifferent in choosing between C and G, as backtracking allows Ag_i to exhaustively move both its choices, if either leads to defeat, rendering strategising unnecessary. Thus:

8. *As both choices seem to be leading to defeat Ag_i may randomly choose between them.*

In the first example the added value of the proposed approach is evident: whereas Ag_i was seemingly in front of a random choice which appeared to be leading

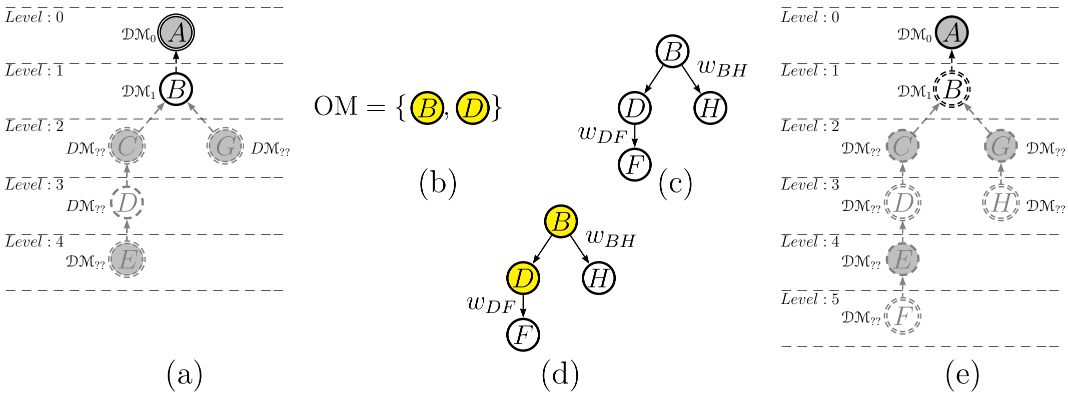


Figure 4.3: Example 5: a) A simulated dialogue based on just the OM, b) Ag_i's OM of Ag_j, c) Ag_i's $\mathcal{R}G$, d) The combination of Ag_i's OM of Ag_j's and its $\mathcal{R}G$.

to the same outcome, the additional information, i.e. the arguments used for augmenting the existing OM, made following just one path an imperative choice.

In the second case though the benefit is not as apparent as one would expect. In contrast to Example 4, the modeller faces the exact opposite situation, as it appears at first that regardless of backtracking all options lead to victory, while it turns out that it is possible for all paths to lead to defeat, deeming choosing between the two possible options, again, a random choice. This example is deliberately provided in order to convey to the reader that the essence of the augmentation mechanism is not so much to provide the means for overcoming what seems to be a decision 'deadlock' when found at a strategic point (e.g. in the previous example, choosing either C or G based on the likelihoods of their possible attackers D and F, defined by weights w_{BD}, w_{DF} if backtracking was not allowed), but mostly in the incorporation of information worth taking into account when building an OM. *How that information will be used and whether it will lead to assisting in decision making or further complicate the decision making process depends on the modeller's objectives, as well as on how one estimates the utility of a certain choice in relation to those objectives.*

For instance, assume that Ag_i's objective is to stall its opponent, or to maintain the dialogue process for as long as possible so as to increase the possibility of encountering new opponent-arguments and thus to collect information about Ag_j aimed to be used in a later more significant dialogue. Then, if F was more likely to be known to Ag_j (i.e. $w_{BD} < w_{DF}$) than H, Ag_i could opt for C.

In general, the essence of the presented examples is to stress how a possible outcome, anticipated through a dialogue simulation, may be altered if additional information is taken into account, and that if this is the case then accounting for that additional information becomes important. Having clarified this, it is worth to further extend Example 5 in order to present a scenario where if one attempts an analysis on the structural information of the arguments which appear in the game, it may be possible to overcome the strategic 'deadlock'.

Example 6 (Example 5 extended) *Let as assume that the following two tables comprise the logical information known to Ag_i and assumed to be known to Ag_j by Ag_i respectively (we are only concerned with the logical premises $\mathcal{K}$ and inference rules $\mathcal{R}$ known to either of them):*

$S_{(i,i)}$	
$\mathcal{K}$	p, y
$\leq'$	$\ldots$
$\mathcal{R}$	$p \Rightarrow q; \sim r \Rightarrow t; \sim f \Rightarrow s; \sim x, y \Rightarrow g$
$\leq'$	$\ldots$
$\mathcal{G}$	$\ldots$

(a)

$S_{(i,j)}$	
$\mathcal{K}$	$\varnothing$
$\leq'$	$\ldots$
$\mathcal{R}$	$\sim s, \sim t \Rightarrow \neg q; \sim g \Rightarrow f; y \Rightarrow r$
$\leq'$	$\ldots$
$\mathcal{G}$	$\ldots$

(b)

Table 4.1: a) Ag_i's knowledge base, b) Ag_i's OM of Ag_j.

Then the arguments which may be instantiated respectively are:

$S_{(i,i)}$	
A	$p; p \Rightarrow q$
C	$\sim f \Rightarrow s$
E	$y; y, \sim x \Rightarrow g$
G	$\sim r \Rightarrow t$
I	y

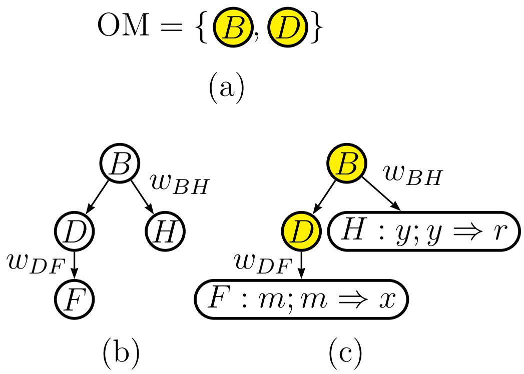
(a)

$S_{(i,j)}$	
B	$\sim s, \sim t \Rightarrow \neg q$
D	$\sim g \Rightarrow f$

(b)

Table 4.2: a) Ag_i's arguments, b) Ag_j's assumed arguments.

Then, given Ag_i's $\mathcal{R}G$ and the relationships between the opponent arguments B,D with other arguments in the $\mathcal{R}G$, particularly H and F, Ag_i may augment



$S_{(i,j)}$	
$\mathcal{K}$	y, m
$\leq'$	$\ldots$
$\mathcal{R}$	$\sim s, \sim t \Rightarrow \neg q; \sim g \Rightarrow f; y \Rightarrow r, m \Rightarrow x$
$\leq'$	$\ldots$
$\mathcal{G}$	$\ldots$

(d)

Figure 4.4: a) Ag_i's OM of Ag_j, b) Ag_i's $\mathcal{R}G$, c) The combination of Ag_i's OM of Ag_j's and its $\mathcal{R}G$, d) Ag_i's augmented OM of Ag_j.

$S_{(i,j)}$ in order to include the logical constituents of H and F. These constituents will be associated with a confidence value[1] *which will result from the relationship weights in the $\mathcal{R}G$.*

Notice that by augmenting $S_{(i,j)}$ through adding premises y and m as well as the defeasible inference rule $m \Rightarrow x$, Ag_i may then assume that Ag_j may be additionally able to instantiate arguments F and H with some likelihood associated with each of them, and which occurs from the confidence values of the logical constituents that compose them. Let us respectively refer to those likelihoods as $P(F)$ and $P(H)$. Let us also accept that these likelihoods are respectively equal to the weights associated with the arguments they are concerned in the $\mathcal{R}G$, i.e.:

$$P(F) = w_{DF}$$

$$P(H) = w_{BH}$$

Ag_i may then extend the simulated dialogue which appears in Figure 4.5a and strategise accordingly. Notice that:

1. *Due to backtracking, at least at an abstract level, opting for either C or G* ***seems*** *pointless.*

[1]At this point we abstain from providing the exact way based on which these values are computed, as we do so in the later sections of this chapter.

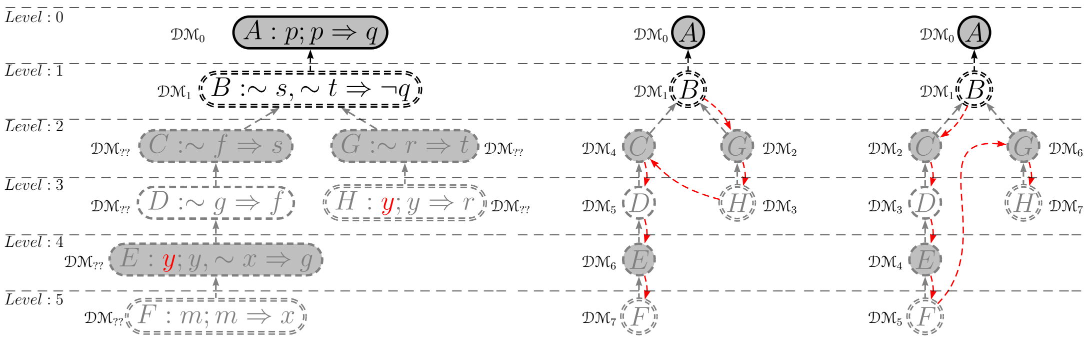


Figure 4.5: Example 5 continued: a) The extended simulated dialogue tree which results from the addition of OM's neighbouring arguments in the $\mathcal{R}G$, b) Dialogue sequence when opting for G, c) Dialogue sequence when opting for C.

2. *This is supported by computing the propagated probability values that characterise how possible it is for Ag_i to face defeat by opting either for C or G, i.e. how likely it is for the dialogue to respectively end at leaf H or F. Let us represent these probabilities as $P_{C \to H}$ and $P_{G \to F}$,then assuming that the intermediate argument D which appears in both of the possible dialogue sequences illustrated in Figures 4.5b & c is known with certainty to the opponent (Ag_j), then these likelihoods are equal to:*

$$P_{C \to H} = P(F) \cdot P(H)$$

$$P_{G \to F} = P(H) \cdot P(F)$$

which are equal.

3. *Let us assume that Ag_i opts for C (Figure 4.5c).*

4. *Let us also assume that argument D is believed to be known to the opponent (Ag_j) with a confidence value equal to 1 (which is the highest possible value).*

5. *Given 2 and 3 it is expected that Ag_i will be forced to move argument E into the game.*

6. *By moving in E Ag_i will be revealing premise y which will be added into its commitment store.*

7. *As we also assume that Ag_j is also aware of the defeasible rule $y \Rightarrow r$ with a confidence value equal to 1, it is then definitely the case that Ag_j will be able to instantiate argument H.*

8. *This means that instantiation of H will no longer be simply probable ($P(H) \neq 0.6$) but will be certain ($P(H) = 1$).*

9. *Accordingly, $P_{C \to H}$ will now be equal to $P_{C \to H} = P(F)$.*

10. *As it now holds that:*

$$P_{C \to H} > P_{G \to F}$$

since:

$$P(F) > P(F)P(H)$$

it is imperative that Ag_i opts for G (Figure 4.5b)

For once more, Example 6 stresses the necessity of accounting for the underlying logic when strategising while it illustrates how doing so combined with information provided from a possible augmentation may prove essential in decision making.

We note that in the above examples we restrain from providing the exact way through which the likelihood values are computed—which define the relationship between certain arguments in the $\mathcal{R}G$ with others instantiated from the contents of an OM. This is a complex issue thoroughly explained in the following sections. At this point it is enough to note that these values depend on the probability values expressed by the weights on the directed edges of a $\mathcal{R}G$.

In relation to the quantification of these probability values, we essentially rely on how often a certain argument follows after another in an agent's history of dialogues, and present two ways based on which this likelihood can be computed. One is presented in this chapter while the other appears in Chapter 5, where we further investigate the modelling aspects of the proposed augmentation process. For instance assume that two arguments A and B are related in a $\mathcal{R}G$ such that:

$$A \xrightarrow{w_{AB}} B$$

we generally assume that w_{AB} is equal to the number of times that A has appeared in dialogues immediately followed by B, divided by the number of times that A has appeared in dialogues. This fraction may then be adjusted to account for additional modelling objectives which will be further discussed in Chapter 5.

Though this appears to be a rather simplistic approach for quantifying the arguments' relationship in a $\mathcal{R}G$, it is nevertheless statistically sound, as it reflects the occurrence ratio of a following argument (e.g. B's occurrence ratio after A). In addition, though this may not be really apparent, through this quantification a modeller manages to inadvertently account for the structure of dialogues when collecting information about its opponent's knowledge. The reader is reminded that in contrast to an agent history an OM ignores the structure of dialogues as it only stores the exchanged logical information. However, by recording this occurrence ratio represented by the arc weights of a $\mathcal{R}G$, through which the likelihood of an argument may then be defined, the modeller accounts for *when* and *how often* certain arguments appear in a game, thus accounting to some extent for the structure of dialogues.

In Chapter 5, the proposed weighting approach is extended to account for the cumulative effect concerned with relationships between indirectly related arguments. Take for example the case of three opponent arguments A, B and C in a relationship graph $\mathcal{R}G$ where we assume relationships between:

$$A \xrightarrow{w_{AB}} B, \quad A \xrightarrow{w_{AC}} C, \quad B \xrightarrow{w_{BC}} C.$$

If we assume that A and C are indirectly related, i.e. that C does follow after A in dialogues but only after B does, which implies a dependency between B and C. Then how should w_{AC} be modelled so as to account for this dependency?

In future we intend to investigate more complex ways for quantifying this likelihood, accounting for contextual factors such as how common certain information is, or whether an agent is part of a certain group having access to shared information, the relationship of certain information with a particular topic etc. Furthermore, an another interesting issue we intend to investigate in the future is the susceptibility of a weighting approach to manipulation. Suppose in our case, that in an attempt to diminish the significance of a relationship between two opponent arguments $A \xrightarrow{w_{AB}} B$, a player can invoke A many times, leaving the modeller anxious to maintain the weighting w_{AB} only with the option of responding B every time A appears.

We should clarify that, though in the full length of this chapter we describe a process which relies on *an abstract representation of arguments*, the incorporated information, which is assumed to be included in the new OM through the augmentation process, concerns the logical information contained in the added arguments. In other words, what will actually become part of the model is the logical constituents of the incorporated arguments. These constituents will also be characterised by a certain confidence value, as is the case with information collected by the other two ICMs. This value will be equal to the likelihood of that information being known to the concerned opponent. Specifically, the constituents of an added argument will inherit the argument's likelihood to represent the modeller's confidence on whether or not the opponent is aware of them. The exact confidence assignation process to the logical constituents of the added arguments as well as how those may be used for defining an arguments' confidence value, after the latter is instantiated from them, is thoroughly discussed in Section 4.5.

To recap, our approach is used when an agent is found at a strategic point in a dialogue. In this case the general approach is to instantiate a dialogue tree, relying on an OM, simulating the possible ways based on which a dialogue game may evolve. Then, each of the possible tree paths is evaluated, based on some utility function, and a choice is made. In a series of steps, the strategising process that utilises the proposed modelling approach, discussed in Examples 4 & 5, is analysed as follows:

1. When found at a strategic point, prior to the simulation of the possible dialogue tree, the concerned OM is first augmented with additional information, based on the relationship between arguments in it and arguments in the $\mathcal{R}G$.

2. The logical constituents of the incorporated arguments resulting from the augmentation process, are associated with a confidence value, one which reflects the confidence value of the arguments instantiated from them, and

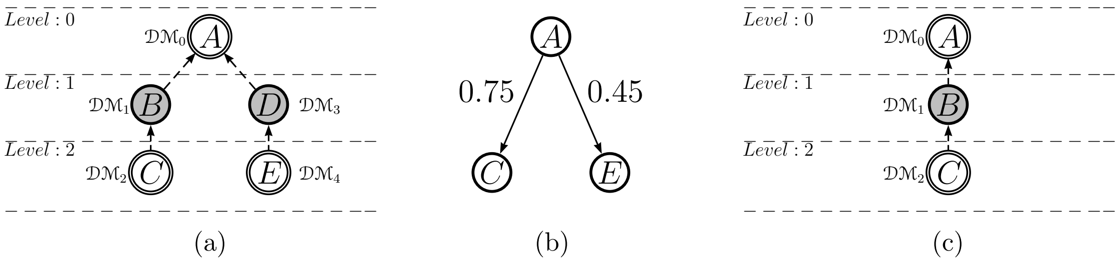


Figure 4.6: a) A dialogue between agents Ag_1 (white) and Ag_2 (grey), b) The resulting $\mathcal{R}G$ induced by Ag_2 , c) A possible dialogue between Ag_3 (white) and Ag_2 (grey).

are added to the OM.

3. After the OM is augmented, the modeller instantiates the new set of arguments that the opponent is assumed to be aware of; associates them with a confidence value that results from the confidence values of its constituents, and; simulates the possible dialogue tree.

4. The utility of all the possible outcomes found at the leaves of this tree, the computation of which accounts for the confidence value assigned to each of the opponent arguments, is then evaluated in order for a choice to be made.

5. The process is repeated until the dialogue is completed.

The above approach concerns an *on the fly* approach taking place during the course of a dialogue game and which accounts for *all* the possible arguments in a $\mathcal{R}G$ which maybe related to arguments instantiated from an OM, assuming that no filtering methods are applied for further distinguishing between subsets of those arguments (e.g. by assuming a likelihood threshold). In the case where such filtering is not applied, the augmentation process will be referred to as an *expansion*, as it relates to a *graph expansion process*, where a sub-graph in a graph is expanded to include all its immediate neighbours (nodes that are at a one hop distance from nodes in the concerned sub-graph).

One may generally act in four different ways:

1. **On the fly expansion:** During the dialogue and when found at a strategic point, the modeller will *expand* its OM with *all* the logical information of

arguments in the $\mathcal{RG}$ that are likely to be related to arguments instantiated from its OM. Thus the modeller will be able to strategise *on the fly* and on the basis that its opponent might be aware of those additional arguments, characterised by a certain level of confidence.

In the trivial case of Example 3 where Ag_2 is assumed to be the modeller, having induced the $\mathcal{RG}$ depicted in Figure 4.6b, would assume that Ag_3 is with some likelihood aware of arguments C and E as soon as Ag_3 moves A into the game, and assuming that these—C and E—are not already part of its OM of Ag_3. Thus, for deciding on whether to deploy argument B or D, Ag_2 would rely on the likelihoods of C and E respectively.

2. **Offline expansion:** At the end of the dialogue, the modeller will *expand* its OM with *all* the logical information of arguments in the $\mathcal{RG}$ that are likely to be related to arguments instantiated from its OM. That is, all the logical information likely to be related to the new arguments used in the terminated dialogue, characterised by a certain confidence value resulting from this likelihood, will now become part of the OM in order to be used in the next time this opponent is encountered.

   In the case of Example 3, assuming that the second dialogue between agents Ag_2 and Ag_3 terminates after arguments $A \leftarrow B \leftarrow C$ are exchanged (Figure 4.6c) and that for some reason related to Ag_2's objectives, though being aware of argument D, Ag_2 opts to not backtrack against A and lose. Then, assuming that Ag_2 is the modeller, at the end of the dialogue Ag_3's OM would be expanded with the inclusion of argument E, given its supporting relationship with A, to be used at a later dialogue.

3. **On the fly augmentation:** During the dialogue and when found at a strategic point the modeller can *attempt* to *augment* its OM with including the logical constituents of arguments likely to be related with arguments instantiated from the OM, but only those with a *high* likelihood, satisfying some predefined threshold. Thus the modeller may strategise on the basis that its opponent might be aware of additional arguments with a confidence higher than a provided threshold which will affect the utility of its choice.

In the trivial case of Example 3 where Ag_2 is assumed to be the modeller, as soon as A is introduced into the game by Ag_3, and assuming a likelihood threshold of 0.5, then, if the relationship likelihoods of arguments C and E are respectively 0.75 and 0.45 (Figure 4.6b), only C would be considered for strategising purposes while E would be ignored.

4. **Offline augmentation:** At the end of the dialogue, the modeller will *attempt* to *augment* its OM with the the logical constituents of arguments likely to be related with arguments instantiated from the OM, but only those with a *high* likelihood, satisfying some predefined threshold. That is, information highly likely to be related to the new arguments used in the terminated dialogue, given satisfaction of a certain threshold, will become part of the OM to be used for the next time this opponent is encountered. The incorporated information will be characterised by a confidence value resulting from how likely they are to be actually aware to the opponent.

   In the case of Example 3, assuming that the second dialogue between agents Ag_2 and Ag_3 terminates after arguments $A \leftarrow B \leftarrow C$ are exchanged (Figure 4.6c) and that for some reason related to Ag_2's objectives, though being aware of argument D, Ag_2 opts to not backtrack against A and lose. Then, assuming that Ag_2 is the modeller, at the end of the dialogue, given a likelihood threshold of 0.5, Ag_3's OM would be augmented with the inclusion of just argument E as its relationship likelihood with A is above the threshold, to be used at a later dialogue.

In essence, augmentation approaches concern approximative solutions which attempt to reduce the search space, while providing equally effective results. As tractability is often a very important practical factor in decision making, the development of approximation methodologies which are both efficient, by reducing complexity, as well as effective, by producing a reliable outcome though sacrificing accuracy, is essential. Obviously, expansion approaches allow one to thoroughly account for all possibilities (e.g. again in relation to Example 3, Ag_2 can account for both the possibility of C as well as E being known to Ag_3 even if one of them has a really low likelihood to be related with an argument instantiated from Ag_3's knowledge base). The added value of opting to expand rather than to augment, is that it is possible for an argument with a low relationship likelihood (which intuitively seems to have no reason to be taken into account) to be part of a dispute which eventually leads to an outcome with the highest utility. Though this may not be obvious at this point, it is thoroughly discussed in Chapter 5 where we analyse how confidence values which occur from the augmentation process can be used for computing the different utility values of the possible simulated paths of a dialogue tree, through the application of a strategy function.

In any case, the objective of this chapter is not differentiate and choose between these four possible options but to provide the technical machinery which allows the application of any of them. After all, an expansion is essentially an augmentation where the likelihood threshold for including an argument in the new OM is set to 0.

We should also note that for the purpose of our work we disregard arguments concerned with logical information in the form of preferences when building a $\mathcal{R}G$. This is because preferences can only be supported by themselves. In other words, preferences that appear after other preferences in a dispute must always be contradictory preferences, and thus the possible options that may follow after a pre-ordering cannot vary. For example, if a pre-ordering $A > B$ is introduced into a game the only thing which can follow after it, is a contradictory pre-ordering $A < B$, which in turn can only be countered by the repetition of $A > B$. In this sense, it is evident that $A > B$ is the only thing that may support itself against $A < B$, which makes modelling the support relationship between preferences useless due to its trivial form.

In addition, for deducing a possible relation of a different kind between the preference orderings moved in dialogues, it is necessary that we are aware of the mechanisms that agents use to form and update their priority-orderings over rules and premises. In general, we assume that agents use generic principles to do so, e.g. the well known *specificity principle*, and the *temporal principle* (which orders newly acquired knowledge over older knowledge). In this work, we leave these mechanisms implicit, as modelling them is out of our research scope. We admit that there is some value in at least including arguments with preferences in a $\mathcal{R}G$ which follow, not after another preference but, immediately after an argument in a dispute thus serving as a different form of support, to previously attacked opponent arguments. However, we choose to leave the provision of a complete augmentation mechanism able to also account for preferences to future work.

4.3.1 Building a Relationship Graph

As noted, intuitively we expect our opponents to be aware of arguments that are likely to follow in a current dialogue, given that they have appeared in previous dialogues and relate to what we currently assume our opponent to know. We assume this likelihood to increase as the relation between the contents of an OM and the arguments external to the model becomes stronger, i.e., as we observe assumed relationships between arguments on the modeller's part being verified by its opponent through a dialogue process. Essentially, the appearance of particular sequences of arguments in dialogues, which relate the two sets, is what validates the modeller's relationship assumptions.

For example, assume that two agents Ag_i and Ag_j, bearing the roles of the proponent and the opponent respectively, engage in a persuasion dialogue, where backtracking is allowed. Let us further assume that the dialogue tree illustrated in Figure 4.7a describes this dialogue. In this case Ag_i and Ag_j introduce arguments $\{A, C, E, G\}$ respectively $\{B, D, F, H\}$. Assume then, that Ag_i engages in another persuasion dialogue with a different agent Ag_m who also happens to counter Ag_i's A with B. It is then possible that Ag_m is *likely* to be aware of arguments D, H or even F. As we have already discussed, this likelihood can be defined with respect to how many times arguments D, H, or F have followed after B in previous dialogues. Our focus in this section is the instantiation of *implied* support relationship between these arguments. Once these relationships are instantiated as links between opponent arguments in a $\mathcal{R}G$, if then Ag_m does indeed put forth arguments D, H and F in the game, then the likelihood of someone knowing D, H and F, contingent that she knows B, should increase.

As discussed, for augmenting an OM ($S_{(i,j)}$), we rely on a $\mathcal{R}G$. For Ag_i a relationship graph ($\mathcal{R}G_i$) is a graph that associates nodes, that represent arguments asserted by an Ag_i's opponents in $\mathcal{H}_i$. This association is represented through weighted directed arcs which represent a support relationship between arguments, while whose weights represent the likelihood of their relationships,

Figure 4.7: a) A dialogue tree $\mathcal{T}$ where the grey and the white nodes concern Ag_1's respectively Ag_2's moves, b) A 1-hop $\mathcal{R}G$ modelling approach, c) A 2-hop $\mathcal{R}G$ modelling approach.

i.e. the extent to which two arguments are related. We assume a $\mathcal{R}G$ to be incrementally built as an agent Ag_i engages in numerous dialogues, being empty at the start, i.e. for $\mathcal{H}_i = \{\varnothing\}$, and constantly updated with newly encountered opponent arguments (OAs). Notice that in the case of the example presented in Figure 4.7(a), assuming that the grey agent is the modeller, OAs (the white's arguments B, D, F and H) can only appear in odd levels of the dialogue tree.

For assigning arcs between these arguments one may rely on how and when an opponent argument appears in a tree, i.e. on the structure of a dialogue tree, since it inherently encodes relationship information between the arguments that appear in it. Though we focus on the support relationship between arguments, one may opt between numerous modelling perspectives for building a $\mathcal{R}G$, focusing on particular aspects of the structure of a dialogue tree that better reflect one's modelling objectives. Different modelling approaches are further discussed in Chapter 5.

Definition 47 (Relationships Graph) *Let $\mathcal{A}^{\mathcal{H}}$ represent the arguments introduced by an agent's opponents in $\mathcal{H}$. Then a $\mathcal{R}G$ is a directed graph $\mathcal{R}G = \{\mathcal{A}^{\mathcal{H}}, R\}$, where $R \subseteq \mathcal{A}^{\mathcal{H}} \times \mathcal{A}^{\mathcal{H}}$ is a set of weighted arcs representing support relationships. We write r_{AB} to denote the arc $(A, B) \in R$, and denote the arc's weight as w_{AB} obtained via a weighting function w, such that $w : R \rightarrow [0, 1]$.*

The modelling approach which we use to capture the support relationship is

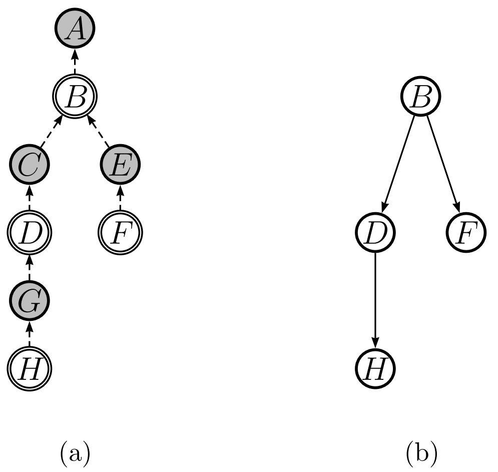


Figure 4.8: a) A dialogue tree $\mathcal{T}$, b) The induced $\mathcal{R}G$ for $\mathcal{T}$.

to assume that two OAs, A and B, are connected in a $\mathcal{R}G$ via an arc, if they are found in the same path of the dialogue tree, i.e. in the same dispute. In addition to appearing in the same dispute, one has to follow at most after a predefined number of levels from the other. For example, consider the case of the dialogue depicted in Figure 4.8a. The induced $\mathcal{R}G$ for opponent arguments that are two levels away from each other appears in Figure 4.8b. Specifically, the opponent arguments B and D are connected in the RG as they both appear in the same dispute line with D following directly after B's attacker in the dispute. This also holds for arguments B and F, as well as for D and H. In contrast arguments D and F appear in different dispute lines and thus are not connected in the $\mathcal{R}G$, while in the case of H and B, though they do appear in the same dispute, their distance exceeds the 2 levels threshold.

At this point we provide two definitions in relation to the support relationships between arguments in order to differentiate between direct and indirect supports, by specifically referring to the first as reinstatement.

Definition 48 (Reinstatement) *Let $d =< \mathcal{DM}_0, \mathcal{DM}_1, \ldots, \mathcal{DM}_n >$ be a dispute in a dialogue tree $\mathcal{T}$ and A, B and C be three arguments that serve as the respective constituents of a subsequence of three dialogue moves in d, such that $< \mathcal{DM}_i, \mathcal{DM}_{i+1}, \mathcal{DM}_{i+2} >$, where $0 \leq i \leq n-2$, we then say that C reinstates A.*

Definition 49 (Indirect support) *Let* $d =< \mathcal{DM}_0, \mathcal{DM}_1, \ldots, \mathcal{DM}_n >$ *be a dispute in a dialogue tree* $\mathcal{T}$ *and* A *and* B *be two arguments that serve as the respective constituents of a subsequence of two dialogue moves in* d*, such that* $< \mathcal{DM}_i, \mathcal{DM}_j >$*, where* $0 \leq i \leq n-2$ *and* $j = i + 2k$*, for* $1 \leq k \leq \lfloor \frac{n-i}{2} \rfloor$*, we then say that* B *indirectly supports* A*.*

We define the distance between two opponent arguments A, B in a dialogue tree as θ_{AB}, and assume a predefined threshold set by the modeller referred to as θ_t for deciding their connectivity in a $\mathcal{R}G$ as follows:

Definition 50 (θ-distance) *Let* $\Delta = \{d_1, \ldots, d_k, \ldots, d_m\}$ *be the disputes of a dialogue tree* $\mathcal{T}$*. Also, let* `level()` *be a function applied on a dialogue move* $\mathcal{DM}_i$ *such that:*

$$\texttt{level}(\mathcal{DM}_i) \rightarrow l_i$$

where l_i *is the level of* $\mathcal{DM}_i$ *in* $\mathcal{T}$*. For every distinct pair of opponent dialogue moves* $\mathcal{DM}_i$ *and* $\mathcal{DM}_j$ *respectively comprising arguments* A *and* B*, then:*

$$\theta_{AB} = \begin{cases} \infty & \textit{if } \mathcal{DM}_i, \mathcal{DM}_j \notin d_k \\ \frac{|l_i - l_j|}{2} & \textit{if } \mathcal{DM}_i, \mathcal{DM}_j \in d_k. \end{cases}$$

Definition 51 (Connectivity Condition) *Let* A *and* B *be two arguments respectively serving as the content of two opponent dialogue moves* $\mathcal{DM}_i$ *and* $\mathcal{DM}_j$ *in a dialogue tree* $\mathcal{T}$*, if:*

- $i < j$*, and;*
- $\theta_{AB} \leq \theta_t$

then $\exists r_{AB} \in \mathcal{R}$*.*

For example, in Figure 4.7a arguments B and D are assumed to be at a $\theta_{BD} = 1$ distance from each other, while B and F are at a $\theta_{BF} = 2$ distance. In addition, Figures 4.7b and 4.7c illustrate two distinct $\mathcal{R}G$s induced from the dialogue tree of Figure 4.7a, for $\theta_t = 1$ and $\theta_t = 2$ respectively. Through modifying the θ_t value one can strengthen or weaken the connectivity, between arguments

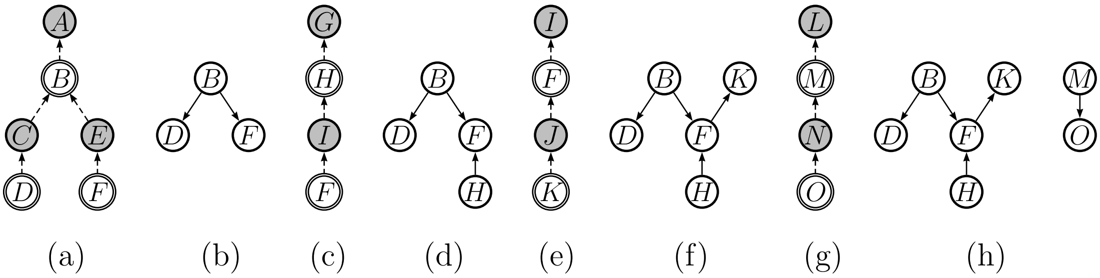


Figure 4.9: a) A dialogue $\mathscr{D}^1$ between Ag_1 (grey) & Ag_2 (white), b) The induced $\mathcal{RG}_1$, c) A dialogue $\mathscr{D}^2$ between Ag_1 (grey) & Ag_3 (white), d) The updated $\mathcal{RG}_1$, e) A dialogue $\mathscr{D}^3$ between Ag_1 (grey) & Ag_2 again (white), f) The updated $\mathcal{RG}_1$, g) A dialogue $\mathscr{D}^4$ between Ag_1 (grey) & Ag_4 (white), h) The final $\mathcal{RG}_1$

in the same path of a $\mathcal{T}$, and correspondingly between arguments in the induced $\mathcal{RG}$. Notice that the direction of the arc that associates two arguments in the $\mathcal{RG}$ is defined by the requirement that $i < j$, without which all relationships would have been reciprocal.

An example of incrementally building a $\theta_t = 1$ $\mathcal{RG}$ after four consecutive persuasion dialogues, where the modeller is Ag_1, is illustrated in Figure 4.9. In this example Ag_1 faces random opponents in Ags, engaging in dialogues which have different topics. The illustrated modelling approach, as well as our modelling hypothesis which relies on the notion of support, are independent of the topic of the dialogues. In other words, it is not necessary that two arguments relate to each other only if they appear in dialogues with the same topic. As far as deducing some properties of the induced graph is concerned, it is easy to see that it is not necessary that $\mathcal{RG}$ is a connected graph. Evidently, arguments M and O are not related in any way with previous OAs and are therefore disconnected from the rest of the graph. Finally, in this simple example, the connectivity between the related arguments is kept to a minimum, both because of the trivial structure of the dialogues, i.e. because the concerned dialogues do not evolve in depths that exceed four levels, but also because of setting θ_t threshold to 1.

Obviously, modelling dialogues with more than, for example, four levels and for a $\theta_t > 1$ would result in a more dense $\mathcal{RG}$ (see Figure 4.7c where for a $\theta_t = 2$ an additional arc is included between arguments B and F). At the same time though, assigning a large value to θ_t raises a *cognitive resources* issue, due to the large volume of information that needs to be stored. Thus, for practical reasons we will generally avoid using large numbers for θ_t. Additionally, it is questionable whether an argument Y, that follows at a large distance after another argument X in the same dispute, actually supports X. A reasonable question raised here is, why not build the $\mathcal{R}G$ only for a $\theta_t = 1$? In other words, what is the added value of inducing $\mathcal{R}G$s with a $\theta_t > 1$ value. Similar questions can be raised in relation to whether arguments should be linked regardless of whether they appear in the same line of dispute or that the relationship between the linked arguments should be reciprocal. However, our intention in this chapter is to provide the basic modelling framework. We address the above questions in the next chapter, where we also propose a more complex way for inferring relationships between arguments, accounting for additional inherent properties between dialogue moves introduced in a dialogue game when inducing a $\mathcal{R}G$.

Lastly, for providing a weight value w_{AB} which will essentially represent the relationship likelihood of an argument A with an argument B, we rely on Definition 52, which is essentially a normalisation that allows us to compute a probability value $Pr(r_{AB}) = w_{AB}$ for arc r_{AB}. We simply count the number of times that an argument A appeared in discrete dialogues followed by B, and we put them against the total number of times that A was used in discrete dialogues. Let us assume for example a weight value $w_{AB} = 0.78$ or 78%. This percentage tells us that out of a total number of 100 times that argument A was used, for 78 of them A was followed by B in the same dispute lines of distinct dialogues, as recorded by a modeller's history of dialogues.

Definition 52 (Weight Assignation) *Let $\mathcal{R}G = \{\mathcal{A}^{\mathcal{H}}, R\}$ be an agent's relationship graph, while arguments A, B are elements of $\mathcal{A}^{\mathcal{H}}$ then:*

$$\texttt{Instances}(\mathcal{H}, A, B) = M_{AB}$$

is a function that returns a number M_{AB} representing the times that argument A was put forth A followed by B in the same disputes in distinct dialogues in $\mathcal{H}$, such that $\theta_{AB} \leq \theta_t$, then:

$$w_{AB} \equiv M_{AB}/M_{A*} \tag{4.3}$$

where $*$ represents the use of any and of no argument at all.

In essence, M_{A*} represents the set of agents that have simply used argument A in distinct dialogues followed by anything or even followed by nothing, i.e. A appears as the leaf-node of a dispute line, while it is evident that:

$$M_{AB} \leq M_{A*}$$

i.e. the denominator will always be at least equal to the enumerator.

We should note that there is a problem with the aforementioned approach. Namely if we consider the example shown in Figure 4.9, based on Definition 52 all the arcs in the induced $\mathcal{RG}$ will initially have a weight value of 1. It is apparent that this value does not represent the real likelihood which relates the arguments at either endpoint of the arc. In order to better approach the *real* likelihood a larger number of dialogues with numerous distinct participants need to be considered. This problem is better known as the *cold start problem* and is encountered in various other contexts as well, (e.g. Lashkari et al. [1994]). A solution to this problem is to initialise the arc weights of an induced $\mathcal{RG}$ with non trivial values representing, in a sense, assumptions based on either expert knowledge or previous experience. However, we leave this for future work.

4.3.2 Relationship augmentation

Having built a $\mathcal{RG}$ an agent Ag_i can then attempt to augment its OM of Ag_j by adding to it the arguments (nodes) that are of n-hop distance in $\mathcal{RG}$ from those contained in that OM and which are most likely to be related with the arguments already in it. As this approach relies on the relationships formed between arguments, for convenience we will be referring to an OM as a set of arguments, those which can be instantiated from the logical information found in an $S_{(i,j)}$, expressed as $\mathcal{A}_{(i,j)}$.

We will also be referring to a possible augmentation of a $S_{(i,j)}$ as $S'_{(i,j)}$, which we assume to be an isomorphic process to the augmentation of $\mathcal{A}_{(i,j)}$ (the incorporation of additional arguments into $\mathcal{A}_{(i,j)}$), which will be referred to as $\mathcal{A}'_{(i,j)}$. For presentation convenience, whenever the subscript $_{(i,j)}$ can follow from the context

$S_{(i,j)}$	
$\mathcal{K}$	$\varnothing$
$\leq'$	$\dots$
$\mathcal{R}$	$\sim s, \sim t \Rightarrow \neg q; \neg g \Rightarrow f; y \Rightarrow r$
$\leq'$	$\dots$
$\mathcal{G}$	$\dots$

(a)

$S'_{(i,j)}$	
$\mathcal{K}$	y, m
$\leq'$	$\dots$
$\mathcal{R}$	$\sim s, \sim t \Rightarrow \neg q; \neg g \Rightarrow f; y \Rightarrow r; m \Rightarrow x$
$\leq'$	$\dots$
$\mathcal{G}$	$\dots$

(b)

Table 4.3: a) The input of function f_{aug}, b) The output of function f_{aug}.

it will be omitted.

Definition 53 (Augmentation) *Let $S_{(i,j)}$ be Ag_i's model of Ag_j's knowledge, then an augmentation is a function f_{aug}:*

$$f_{aug}(S_{(i,j)}) \mapsto S'_{(i,j)}$$

through which new information is incorporated into $S_{(i,j)}$.

For example, let us recall the opponent model $S_{(i,j)}$ presented in Example 6 in Table 4.3b, which represents Ag_i's assumption of Ag_j's knowledge. $S_{(i,j)}$'s augmentation $S'_{(i,j)}$ which results from the relationship between the arguments B and D instantiated from $S_{(i,j)}$ with arguments F and H in Ag_i's $\mathcal{R}G$, appears in Figure 4.4b. Both tables appear in Tables 4.3a, 4.3b as the input respectively output of f_{aug}, depicting the augmentation process in an abstract way.

As the augmentation process focusses on the relationships between the arguments instantiated from an OM and arguments in a $\mathcal{R}G$, rather than between their logical constituents, henceforth we will be referring to an augmented OM only as $\mathcal{A}'_{(i,j)}$ and will abstain from using its analytic form ($S'_{(i,j)}$). Thus we may at times use a more high level form of f_{aug}, assuming its application on $\mathcal{A}_{(i,j)}$ as opposed to $S_{(i,j)}$, e.g.:

$$f_{aug}(\mathcal{A}_{(i,j)}) \mapsto \mathcal{A}'_{(i,j)}.$$

We stress though that $\mathcal{A}'_{(i,j)}$ does not necessarily encapsulate all arguments that may be instantiated from $S'_{(i,j)}$[1].

[1]We remind the reader that new arguments may be instantiated once new logical information

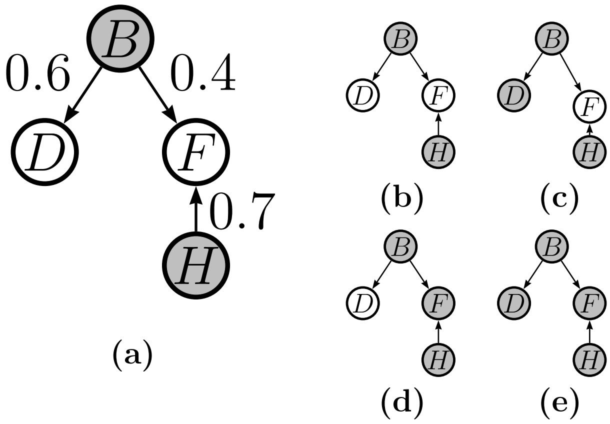


Figure 4.10: a) $\mathcal{RG}_1$, b), c), d) & e) Possible augmentations $\mathcal{A}'_{\varnothing}$, $\mathcal{A}'_D$, $\mathcal{A}'_F$, & $\mathcal{A}'_{DF}$ respectively.

In order to understand the complexity of the proposed approach we examine a trivial scenario for the augmentation of an OM. Let us assume a $\mathcal{RG}_1$ induced by Ag_1 as it is illustrated in Figure 4.10a, where for the purpose of this example we assume that the weights on the arcs have received their values after numerous dialogues. Let us assume that based on Ag_1's OM of Ag_4, Ag_1 believes that Ag_4 is aware of two arguments $\mathcal{A}_{(1,4)} = \{B, H\}$ (the grey nodes in Figure 4.10a). Based on $\mathcal{RG}_1$ the possible augmentations of $\mathcal{A}_{(1,4)}$ are $\mathcal{A}'_{\varnothing}$, $\mathcal{A}'_D$, $\mathcal{A}'_F$, and $\mathcal{A}'_{DF}$ as they respectively appear in Figures 4.10b,c,d & e. In this example, instead of enumerating the possible augmentations with numerical subscripts, we list the added arguments as subscripts to make the example easier to follow.

For augmenting $\mathcal{A}_{(1,4)}$ the modeller Ag_1 has to choose between these augmentations the one with the highest likelihood. In the following example we illustrate in a step by step process how these likelihoods are computed.

Example 7 *Assume we want to calculate the likelihood of augmentation $\mathcal{A}_{(1,4)}$ to $\mathcal{A}'_F$. In this simple example the likelihood of including argument F is the likelihood of including the* ***in****-coming arcs to F*[1]*. These are r_{HF} and r_{BF}, which means that*

is incorporated into an existing $S_{(i,j)}$.

[1]Since the weights on the arcs indicate the frequency with which the connected arguments follow each other in Ag_1's history.

F can be included either if r_{HF} or r_{BF} is chosen, or if both are chosen. Therefore:

$$\begin{aligned} Pr(F) &= Pr(r_{HF} \cup r_{BF}) \\ &= Pr(r_{HF}) + Pr(r_{BF}) - Pr(r_{BF} \cap r_{HF}) \\ &= w_{HF} + w_{BF} - w_{BF} \cdot w_{HF} = 0.82 \end{aligned}$$

The probability of inducing $\mathcal{A}'_F$ is the probability of including argument F and not including D which is:

$$Pr(\mathcal{A}'_F) = Pr(F)(1 - Pr(D)) = Pr(F)(1 - w_{BD}) = 0.328.$$

Similar computations need to take place for calculating the likelihoods of the rest of the possible augmentations ($\mathcal{A}'_\varnothing$,$\mathcal{A}'_D$, and $\mathcal{A}'_{DF}$). For providing a general formula for computing the likelihood of a possible augmentation, we rely on basic graph theory notation with respect to a node X in a graph $\mathcal{R}G$, such as:

- **`in`-degree of X ($d^+(X)$):** The number of arcs adjacent to X which end at X.
- **`out`-degree of X ($d^-(X)$):** The number of arcs adjacent to X which begin from X.
- **degree of X ($d(X)$):** The total number of adjacent arcs that either end at or begin from X:
$$d(X) = d^+(X) + d^-(X)$$
- **neighbouring nodes of X ($N(X)$):** The union product of the set of nodes adjacent to X with `in`-bound arcs $N^+(X)$, and with `out`-bound arcs $N^-(X)$:
$$N(X) = N^+(X) \cup N^-(X) \quad \textit{where} \quad |N(X)| = d(X)$$
- **adjacent arcs of X ($R(X)$):** The union product of the set of `in`-bound arcs $R^+(X)$, and `out`-bound arcs $R^-(X)$ adjacent to X:
$$R(X) = R^+(X) \cup R^-(X) \quad \textit{where} \quad |R(X)| = d(X)$$

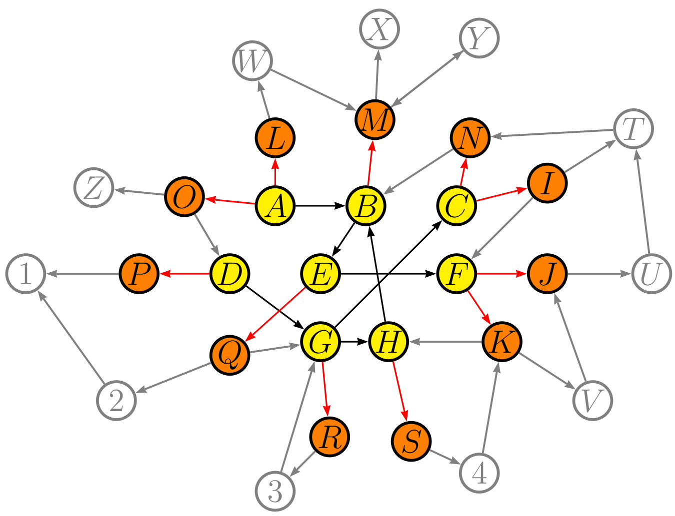


Figure 4.11: A $\mathcal{R}G$ where the yellow nodes represent set $\mathcal{A}$, the orange nodes represent set $N_{\mathcal{A}}$, and the red arcs represent set $R_{\mathcal{A}}$ (Please refer to this figure in colour).

In addition, assuming a set of arguments $\mathcal{A}$ (the yellow nodes in Figure 4.11), we define two sets $N_{\mathcal{A}}$ and $R_{\mathcal{A}}$ where:

- **neighbouring nodes of $\mathcal{A}$ ($N_{\mathcal{A}}$):** Represents the neighbours of the nodes in $\mathcal{A}$ (the orange nodes in Figure 4.11), and is formed from the union of the neighbours of every node X in $\mathcal{A}$, excluding those that are already in $\mathcal{A}$:

$$N_{\mathcal{A}} = \bigcup_{X \in \mathcal{A}} N(X) - \mathcal{A}$$

- **adjacent arcs of $\mathcal{A}$ ($R_{\mathcal{A}}$):** Is the adjacent arcs of the nodes in $\mathcal{A}$ (the red arcs in Figure 4.11) and is equal to the adjacent arcs of every element X in $\mathcal{A}$, excluding those that connect with arguments already in $\mathcal{A}$:

$$R_{\mathcal{A}} = \bigcup_{X \in \mathcal{A}} R(X) - \{r_{XY} Y \notin \mathcal{A}\}.$$

We note that for $\mathcal{A}_{(i,j)}$ it reasonably holds that $\mathcal{A}_{(i,j)}$ is a subset of the opponent arguments instantiated from Ag_i's overall history $\mathcal{A}^{\mathcal{H}_i}$, i.e. $\mathcal{A}_{(i,j)} \subseteq \mathcal{A}^{\mathcal{H}_i}$. For convenience we will hence refer to a $\mathcal{A}_{(i,j)}$ as $\mathcal{A}$ and to its augmentation as $\mathcal{A}'$.

Given these, we can now provide the following definitions:

Definition 54 (Possible augmentations) *Let $P = \{\mathcal{A}'_0, \mathcal{A}'_1, \ldots, \}$ be the set of the possible augmentations of $\mathcal{A}$. Then the number of all of $\mathcal{A}$'s distinct possible augmentations $|P|$ with respect to the neighbouring nodes that are of 1-hop distance from $\mathcal{A}$, is the sum of the k-combinations of the elements of $N_{\mathcal{A}}$, for $k = 0, \ldots, |N_{\mathcal{A}}|$:*

$$|P| = \sum_{k=0}^{|N_{\mathcal{A}}|} \binom{|N_{\mathcal{A}}|}{k} = 2^{|N_{\mathcal{A}}|} \tag{4.4}$$

each of which is given by the binomial coefficient $\binom{|N_{\mathcal{A}}|}{k}$ where:

$$\binom{|N_{\mathcal{A}}|}{k} = \frac{|N_{\mathcal{A}}|!}{k!(|N_{\mathcal{A}}| - k)!}.$$

In the case of Example 7, since $|N(\mathcal{A}_{(1,4)})| = 2$ (arguments D and F) the possible augmentations of $\mathcal{A}_{(1,4)}$ is:

$$|P| = \sum_{k=0}^{2} \binom{2}{k} = \binom{2}{0} + \binom{2}{1} + \binom{2}{2} = 1 + 2 + 1 = 4.$$

For providing the general formula for computing the likelihood of a possible augmentation $\mathcal{A}'$ one has to simply compute the product of the probabilities of every neighbouring argument Y in $N(\mathcal{A})$ assumed to be in the augmented set ($\mathcal{A}'$), multiplied by the product of the probabilities of every neighbouring argument Y in $N(\mathcal{A})$ assumed *not* to be in the augmented set ($\mathcal{A}'$). In the simple case of Example 7 this was reflected by multiplying the probability of including argument F ($Pr(F)$) and not including argument D ($1 - Pr(D)$).

Definition 55 (Likelihood formula) *Let $\mathcal{A}$ be a set of arguments instantiated from an opponent model S, and $N(\mathcal{A})$ the neighbouring nodes of $\mathcal{A}$. Then, the general formula for computing the likelihood of a possible augmentation $\mathcal{A}'$ with respect to each of the neighbouring nodes (arguments) of a $\mathcal{A}$, i.e. for every*

$Y, Z \in N(\mathcal{A})$ *where* $Y \neq Z$, *is:*

$$Pr(\mathcal{A}') = \prod_{Y \in \mathcal{A}'} Pr(Y) \prod_{Z \notin \mathcal{A}'} (1 - Pr(Z)) \tag{4.5}$$

where letting $R^+(Y)$ *be the* ***in****-bound arcs of* Y*, then:*

$$Pr(Y) = \sum_{r \in R^+(Y)} Pr(r) - \prod_{r \in R^+(Y)} r. \tag{4.6}$$

In essence, Equation 4.6 is the union product of the probabilities associated with each of the the **in**-bound arcs of Y, and is computed in accordance to the inclusion exclusion princible for probability, the representation of which requires that we index the elements of the set of the in-bound arcs R^+ of a neighbouring argument Y, such that $R^+(Y) = \{r_1, r_2, \ldots, r_n\}$:

$$\begin{aligned} Pr(Y) &= Pr(\bigcup_{r \in R^+(Y)} r) \\ &= \sum_{i=1}^{n} Pr(r_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Pr(r_i \bigcap r_j) + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} Pr(r_i \bigcap r_j \bigcap r_k) \ldots . \end{aligned}$$

where $r_i, r_j, r_k, \ldots \in R^+(Y)$.

Lastly, since the likelihood of each possible augmentation should define a distribution of likelihoods then it must also hold that:

$$\sum_{\mathcal{A}' \in P} Pr(\mathcal{A}') = 1. \tag{4.7}$$

At this point it is essential to mention that Equation 4.5 heavily relies on knowing the likelihood of each argument. This is because, while the number of possible augmentations of a set $\mathcal{A}$ is *exponential* in the size of the adjacent nodes ($O(2^{N_{\mathcal{A}}})$), determining the augmentation with the highest likelihood is equivalent to simply determining the arguments with the highest likelihoods, or more precisely those that have a likelihood higher than 0.5. As counter-intuitive as this may be, if:

$$Pr(X) > 1 - Pr(X). \tag{4.8}$$

holds for all included arguments of a possible augmentation $\mathcal{A}'$, then Equation 4.5 obtains its highest possible value. In other words, if an argument has a likelihood above 0.5, which therefore serves as a selection threshold ($\theta_S = 0.5$) then it will definitely be part of the augmentation with the highest likelihood.

Particularly, threshold θ_S results from solving Equation 4.8 with respect to $Pr(X)$:

$$\begin{aligned} & Pr(X) > 1 - Pr(X) \\ \Leftrightarrow & 2 \cdot Pr(X) > 1 \\ \Leftrightarrow & Pr(X) > \frac{1}{2}. \end{aligned}$$

For example, notice that in the case of the possible augmentations $P = \{\mathcal{A}'_{\varnothing}, \mathcal{A}'_D, \mathcal{A}'_F, \mathcal{A}'_{DF}\}$ that appear in Figure 4.10, the augmentation with the highest likelihood contains, indeed, two arguments with a likelihood value above 0.5, and that is augmentation $\mathcal{A}_{DF}$.

After computing the distinct likelihoods of arguments D and F, the likelihood of each of the possible augmentations is computed as follows:

$$\begin{aligned} Pr(F) &= Pr(r_{HF} \cup r_{BF}) \\ &= Pr(r_{HF}) + Pr(r_{BF}) - Pr(r_{BF} \cap r_{HF}) \\ &= w_{HF} + w_{BF} - w_{BF} \cdot w_{HF} = 0.82 \\ Pr(D) &= Pr(r_{BD}) = w_{BD} = 0.6 \\ Pr(\mathcal{A}'_{\varnothing}) &= (1 - Pr(F)) \cdot (1 - Pr(D)) = (1 - 0.6) \cdot (1 - 0.82) = 0.072 \\ Pr(\mathcal{A}'_D) &= Pr(D) \cdot (1 - Pr(F)) = 0.6 \cdot (1 - 0.82) = 0.10.8 \\ Pr(\mathcal{A}'_F) &= Pr(F) \cdot (1 - Pr(D)) = 0.82 \cdot (1 - 0.6) = 0.328 \\ Pr(\mathcal{A}'_{DF}) &= Pr(F) \cdot Pr(D) = 0.82 \cdot 0.6 = 0.492. \end{aligned}$$

This is further supported by the following proposition:

Proposition 3 *Assume a $\mathcal{RG}$, an opponent model $\mathcal{A}$ and an augmentation $\mathcal{A}'$ of $\mathcal{A}$ such that $Pr(\mathcal{A}')$ is the highest likelihood. Assume that $\mathcal{RG}$ is expanded to include an additional argument X such that X is a neighbouring node of the arguments in $\mathcal{A}$. If X is included in $\mathcal{A}'$ then $\mathcal{A}'$ remains the augmentation with* *the highest likelihood if:*

$$Pr(X) > 1 - Pr(X).$$

Proof We prove this by contradiction. Let us assume that it holds that $Pr(X) < 1 - Pr(X)$, while $Pr(\mathcal{A}' \cup X) = Pr(\mathcal{A}') \cdot Pr(X)$ is the highest likelihood. Then it must also hold that:

$$\begin{aligned} &Pr(\mathcal{A}') \cdot Pr(X) > Pr(\mathcal{A}') \cdot (1 - Pr(X)) \\ \Leftrightarrow &Pr(X) > 1 - Pr(X) \end{aligned}$$

The latter obviously contradicts with our original assumption.

Finally, given that $\mathcal{A}'_{DF}$ is the augmentation with the highest likelihood, its arguments (D and F) will be chosen to be incorporated into $\mathcal{A}$, while $Pr(\mathcal{A}'_{DF})$ will be used to denote the confidence value c of the logical constituents of those arguments which will be incorporated into $S'_{(1,4)}$, as defined in Definition 46(c).

4.4 The Monte-Carlo Simulation

A drawback of the proposed approach is that, the fastest (known) algorithms for calculating the probability of Equation 4.6 have run-time exponential in their input size. This makes the approach practically intractable. However, drawing inspiration from the work of Li et al. [2011] we rely on an approximate approach for computing these likelihoods based on a Monte-Carlo simulation. For the case of the 1-hop augmentation—augmenting with the direct neighbours—we know that the number of possible augmentations is exponential on the size of $N_{\mathcal{A}}$ (Equation 4.4). However, it is possible to deduce a high likelihood augmentation in linear time, provided we know $Pr(Y)$ for $\forall Y \in N_{\mathcal{A}}$. Also, $Pr(Y)$ is calculated through Equation 4.6 using the inclusion-exclusion principle found in basic algorithm textbooks such as Knuth [1997]. But computation of $Pr(Y)$ also has exponential run-time on the in-degree to calculate. We therefore proceed to sample for $Pr(Y)$, by describing a *Monte-Carlo* method to sample for high likelihood arguments, which we will include in our 1-hop augmentation of a set $\mathcal{A}$. We will generally refer to the arguments in $\mathcal{A}$ as augmentation nodes.

In essence, Monte-Carlo methods rely on sampling, repeatedly and randomly, through running a series of simulations for the purpose of heuristically calculating probability values, the theoretical computation of which is very expensive. A simple and intuitive example which will assist in explaining the complex aspects of the proposed sampling procedure analysed in this section, is an experiment concerned with approximately computing the probability of getting at least one 6 when throwing two dice.

Obviously in a perfect world one would assume that this probability is equal to the probability of getting a 6 with the first die $1/6$, plus the probability of getting a 6 with the second die $1/6$, minus their joint probability $1/36$, ergo $11/36$. This is the theoretical probability value. However this is based on various assumptions which may not seem so apparent. For example, that the dice are unbiased—with their weights evenly distributed on all six sides, no worn edges, etc—or that the surface on which they are thrown is flat and so on. Accounting for these details makes computing this probability a much more complex task.

A solution could be given by performing the following experiment. That is to throw the dice for a large number of times, i.e. 6000 times, and record the results referred to as *sample*. Through recording the results of this sampling method and assuming that the method is well defined so as to guarantee convergence, i.e. that the experimental probability will indeed be an approximation of the theoretical probability, one may acquire an approximation of the desired theoretical values. In our case this can be done in a tractable and quick way, since directly computing the theoretical probability values of the possible augmentations would be much more expensive.

We propose a similar method for our simulation. There are two ways we can move from here in order to define our sampling approach:

1. Obtain a sample of n augmentations $\mathcal{A}'$ and infer their likelihood by their frequency of occurrence.

2. Obtain a sample of n augmentations $\mathcal{A}'$ and infer the likelihood of each of the nodes (arguments) by the frequency in which they were included in

augmentations.

We proceed with the second approach, simply because, as mentioned above, the number of possible augmentations are exponential in the number of possible nodes that can be included. Such a large distribution size means that probabilities will be very low and almost uniform if we assume the first sampling approach. In this case even the slightest sampling error would be unacceptable. In addition, due to Proposition 3 determining the argument likelihood is in essence all that is required for determining the highest likelihood augmentation. Thus we are not interested in determining the likelihoods of all augmentations but instead only of a single high likelihood augmentation. In other words, since, based on Proposition 3, this particular augmentation is obtained by including each argument Y with probability $Pr(Y) > 0.5$, our simulation is a process which attempts to do exactly that: approximately determine whether an argument $Y \in N_{\mathcal{A}}$ should be part of the augmentation with the highest likelihood, according to whether its inclusion frequency in the sampled augmentations exceeds the threshold of $\theta_S = 0.5$.

Abstractly described, the method we propose is described in the following series of steps:

1. Assume a $\mathcal{R}G$= $\{\mathcal{A}^{\mathcal{H}}, R\}$ and a set $\mathcal{A}$.
2. We begin with an initial augmentation $\mathcal{A}' = \mathcal{A}$.
3. For each argument $Y \in N_{\mathcal{A}}$, if $\exists r_{XY} \in R$ where $X \in \mathcal{A}$, we accept r_{XY} if w_{XY} is greater that a random value *random* between 0 and 1.
4. If an arc r_{XY} is accepted, we then add Y to $\mathcal{A}'$ and increase a counter k_Y by 1, which monitors the number of times that Y was included in an augmentation $\mathcal{A}'$.
5. At the end of the process, $\mathcal{A}'$ contains a possible 1-hop augmentation of $\mathcal{A}$.
6. Steps 1 to 5 are repeated for a number of times equal to n which is sufficient for guaranteeing convergence to the theoretical probability values of including each of the elements of $N_{\mathcal{A}}$ in $\mathcal{A}'$.

ALGORITHM OneHopAugmentation($\mathcal{R}G, \delta, \epsilon, \mathcal{A}$)
Ensure: $n, k_Y \in K$ for all Y which were sampled

$K \leftarrow \emptyset$
$n \leftarrow z_\delta^2 \frac{1}{4\epsilon^2}$
for $i = 1$ to n **do**
  $\mathcal{A}' \leftarrow \mathcal{A}$
  for all $r_{XY} \in R_\mathcal{A}$ **do**
    random $\leftarrow$ Random$(0 - 1)$
    if $w_{XY} \geq$ random **then**
      if $Y \notin \mathcal{A}'$ **then**
        $\mathcal{A}' \leftarrow \mathcal{A}' \cup \{Y\}$
        if $k_Y \in K$ **then**
          $k_Y \leftarrow k_Y + 1$
        else
          $k_Y \leftarrow 1$
          $K \leftarrow K \cup \{k_Y\}$
        end if
      end if
    end if
  end for
end for

Algorithm : 4.1: Monte-Carlo simulation

7. Finally, we estimate the likelihood of each of the arguments in $N_\mathcal{A}$ (by dividing their distinct observations k_Y by n) and infer a probability distribution of arguments[1].

Steps 1 to 6 are analytically described in Algorithm 4.1.

In the case of the dice experiment, inferring a probability distribution of arguments would be equivalent with counting the number of times we got a 6, a 5, a 4 and so on, on at least one die—when throwing both—divided by 6000 to infer the frequency of that event. Using any selection policy we wish, we can construct a set of arguments we believe that the opponent is likely to know. This policy can be a specific cut-off, e.g. we only include an argument Y with $Pr(Y) > \theta_S$, where θ_S can be set to any value, e.g. 70% or 60%. For the purpose of our work

[1]A probability distribution assigns a probability to each measurable subset of the possible outcomes of a random experiment. In our case these are the arguments considered for the augmentation.

we assume θ_S to be set to 50% in order to reflect the properties of Proposition 3.

Let us further analyse the proposed algorithm, beginning with explaining the variables used:

$\mathcal{R}G$: The relationships graph.

ϵ: The acceptable error bound for an estimate.

δ: A proportion of the experiments which produce estimates with error values not within the $\epsilon\pm$ bounds.

z_δ: Refers to $z_{1-\frac{\delta}{2}}$ which is the normal distribution quantile function.

$\mathcal{A}$: The set of arguments instantiated from an OM.

$\mathcal{A}'$: A possible augmentation of $\mathcal{A}$.

K: The set of observations of arguments included in the sampled augmentations.

k_Y: The number of times that argument Y was observed in a sampled augmentation.

The first step of the algorithm is to initialise set K as an empty set ($\emptyset$).

$K \leftarrow \emptyset$

Following on n is assigned with $z_\delta^2 \frac{1}{4\epsilon^2}$ which the minimum number for which we must repeat the experiment in order to guarantee convergence to the theoretical probability values of each of the neighbouring arguments. We elaborate more on how this value is computed in Section 4.4.1.

$n \leftarrow z_\delta^2 \frac{1}{4\epsilon^2}$

Next, we inititate a **for** loop repeated for n times, in which the first step is to assign to $\mathcal{A}'$ the content of $\mathcal{A}$, i.e. at this point the augmented set contains no additional arguments.

for $i = 1$ to n **do**
    $\mathcal{A}' \leftarrow \mathcal{A}$
    ...

```
end for
```

Then, for every arc r_{XY} in $R_{\mathcal{A}}$ where X in $\mathcal{A}$ and Y in $N_{\mathcal{A}}$ we produce a random number between 0 and 1:

```
for all r_XY ∈ R_A do
    random ← Random(0 − 1)
    ...
end for
```

and if the arc's corresponding weight value w_{XY} is greater or equal to that number:

```
if w_XY ≥ random then
    ...
end if
```

then if argument (Y) is not already part of $\mathcal{A}'$, it is included in $\mathcal{A}'$:

```
if Y ∉ A' then
    A' ← A' ∪ {Y}
    ...
end if
```

and if a k_Y, the variable responsible for recording the number of times that Y is included in an augmentation, is already an element of K, then its existing value is simply increased by 1, else k_Y is initialised with the value of 1 and is added into K:

```
if k_Y ∈ K then
    k_Y ← k_Y + 1
else
    k_Y ← 1
    K ← K ∪ {k_Y}
end if
```

In essence, according to Algorithm 4.1, the probability of accepting an argument Y is assumed to be equal to the probability of accepting the arc in which it participates as the end node. In turn, the probability of accepting an arc $Pr(r_{XY})$ is equal to the arc's weight value $Pr(r_{XY}) = w_{XY}$. We also note that we assume the events of accepting r_{ij} and r_{km}, where $r_{ij} \neq r_{km}$ to be independent but not mutually exclusive.

Equivalently, in the case of the dice experiment, this means that getting a 6 on one die is *independent* of what we get on the other and at the same time it does not mean that if we get a 6 on one die we wont get a 6 on the other, i.e. *not mutually exclusive*. Correspondingly, the probability of accepting of r_{ij} ($Pr(r_{ij}) = w_{ij}$) is independent of the probability of accepting r_{km} ($Pr(r_{km}) = w_{km}$), while both can also happen at the same time ($Pr(r_{ij} \cap r_{km})$).

Therefore, for computing the probability of accepting both $r_{ij} \cup r_{km}$ we rely on the following formula:

$$Pr(r_{ij} \cup r_{km}) = Pr(r_{ij}) + Pr(r_{km}) - Pr(r_{ij} \cap r_{km})$$

This means that the probability for a node (argument) to be added in $\mathcal{A}'$ follows Equation 4.6. Consequently the probability of obtaining a specific augmentation $\mathcal{A}'$ follows Equation 4.5, and therefore the described procedure essentially samples from the probability distribution of augmentations. In other words, though it seems that the proposed Monte-Carlo simulation, which relies on Algorithm 4.1, behaves differently in relation to how the theoretical values that characterise the likelihood of a certain augmentation are decided; since it deals with likelihoods that characterise arguments rather than augmentations, the fact that the probability of a node being added in $\mathcal{A}'$ follows Equation 4.6 serves as proof that it essentially does the same thing—it samples from the probability distribution of augmentations.

Assuming a sampling procedure which generates a number of n samples using the described method, then each node Y is included with probability equal to $Pr(Y)$. Thus after n independent identically distributed (*i.i.d.*) samples, a proportion of $n \cdot Pr(Y)$ will contain argument Y. We define:

$$k_Y = \sum_{i=1}^{n} I(Y \in \mathcal{A}'_i)$$

where I is the indicator function, taking the value of 1 if the predicate within is satisfied and $\mathcal{A}'_i$ is the set of nodes contained in the i-th sampled augmentation. In the dice experiment the value of the k_Y indicator would be the number of times we rolled either $1:6$, $2:6$, $3:6$, $4:6$, $5:6$, or $6:6$. In our case k_Y is used to denote the number of times we sampled Y.

4.4.1 Sampling accuracy & Experimental results

The development of any simulation for the approximative computation of certain theoretical values needs to be characterised by certain properties which can guarantee that the concerned simulation will eventually converge to the theoretical results, while as well that it is feasible. These properties are related with the notions of:

ϵ: Experimental error of an argument's likelihood to be included in $\mathcal{A}'$, which, at least abstractly, represents the deviation of the calculated probability approximation from the actual theoretical value;

$\mathbf{E}\{k_y\}$: The expected number of observations of a certain argument Y;

$\mathbf{Var}\{k_y\}$: The variance of each of these observations, which is a non-negative value that measures how far a set of numbers is spread out[1];

δ: A value, produced through mathematical analysis, which represents our confidence that ϵ will be within the bounds of a certain threshold, and;

$\hat{Pr}(Y)$: The expected probability of accepting a Y in a $\mathcal{A}'$.

We note that for the computation of the experimental error (ϵ) it is necessary that we do compute the actual theoretical probabilities for all the arguments we consider in our experiments.

Generally, the feasibility of using a Monte-Carlo simulation is based on:

- whether or not the sampling procedure yields a sample from the distribution in question;
- if each sample is independent of previous samples;

[1] A small variance indicates that the data tend to be very close to the mean (expected value) and hence to each other, while a high variance indicates that the data are very spread out around the mean and from each other.

- if running the simulation is much more efficient and tractable than exhaustively computing the theoretical probabilities; and
- whether it accurately computes the underlying probability distribution.

Based on these, the augmentation procedure which requires determining the argument likelihoods makes for an ideal candidate for the use of the Monte-Carlo simulation as described in Algorithm 4.1. The proof for quick convergence is presented in this section.

Let us again consider the dice experiment. Essentially, the principle is the same and at the end of each experiment we will obtain an estimated probability for getting at least a 6 on one of the dice. Ideally we want this estimate to be 11/36. However in practice we get a result which will be reasonably close to that. In this case the ϵ would be an error value meaning that we would accept an estimate which is $\pm\epsilon$ away from the real value (11/36). At the same time, we would like in the results of this estimate to be within ϵ bounds for most of the times we perform the experiment, e.g. for a proportion of δ of the experiments we might get a value outside these bounds, but for the rest of the experiment that value will be within these bounds.

The expected number of augmentation samples which contain Y follows a *binomial distribution*, as it is the case with the dice experiment. Generally, the binomial distribution is the discrete probability distribution of the number of successes in a sequence of n *independent* success/failure experiments, each of which yields success with a certain probability. Our trivial dice experiment is one such experiment. This distribution is frequently used to model the number of successes in a sample of size n, i.e. in the dice experiment, how many times we expect to observe a 6 after n rolls. Equivalently in the case of our experiment that is how many times we expect an argument Y to be chosen after n tries.

Due to the law of large numbers and the fact that each sample is *i.i.d.*, it holds that the expected number of observations k_Y, of any given node Y for a number of tries n is:

$$\mathbf{E}\{k_Y\} = n \cdot Pr(Y) \tag{4.9}$$

and the variance is:

$$\mathbf{Var}\{k_Y\} = n \cdot Pr(Y) \cdot (1 - Pr(Y)) \tag{4.10}$$

Note that the above equations concern a *single* experiment. This defines a multinomial distribution over the set of nodes, i.e. a binomial distribution for every neighbouring argument. Therefore we can estimate the probability of accepting Y as follows:

$$\hat{Pr}(Y) = \frac{k_Y}{n} \tag{4.11}$$

which directly follows from Equation 4.9 and where $\hat{Pr}(Y)$ denotes our estimates of $Pr(Y)$. In other words, the estimate of the likelihood of a certain argument Y would be the frequency of which we sampled it.

To recap, the procedure described here is a method to sample for the inclusion-exclusion probability as an alternative of exhaustively calculating these probabilities, which is of exponential complexity in the number of arcs considered. As this is an approximative solution we expect that the sampled probability will be characterised by some error which we intend to bound, deviating from the actual theoretical values. Bounding that error can be achieved by calculating the number of necessary samples n we need to collect prior to computing these probabilities. The number of samples n required to achieve an accuracy an error ϵ with a confidence δ for the case of the augmentation sampling, is given in the following theorem:

Theorem 6 *The Monte-Carlo approach to sample the probability distribution of the nodes in a set $N_{\mathcal{A}}$ is at least ϵ close to convergence, with probability at least δ after:*

$$n = z_\delta^2 \frac{1}{4\epsilon^2} \quad \textit{samples} \tag{4.12}$$

Proof Assume we sample n points from $\mathcal{A}^{\mathcal{H}_i}$. This would yield a sample K. We now need to determine how far the empirical observations k_Y are from the expected $\mathbf{E}\{k_Y\}$. We use the normal approximation interval which can be found in various textbooks, e.g. Mood [1963]. Since we know that for each argument in

$N_{\mathcal{A}}$ the probability to observe it in a given sample follows a binomial distribution, the normal approximation can be expressed as follows:

$$Pr\left(|\hat{Pr}(Y) - Pr(Y)| \geq z_{1-\frac{\delta}{2}}\sqrt{\frac{\hat{Pr}(Y)(1-\hat{Pr}(Y))}{n}}\right) \leq \delta \tag{4.13}$$

where $z_{1-\frac{\delta}{2}}$ is the normal distribution quantile function (the inverse of the CDF[1]) which gives the value for which a standard normal random variable Y has the probability of exactly δ to fall inside the range $(-\infty, z_{1-\frac{\delta}{2}}]$.

We set:

$$\epsilon = z_{1-\frac{\delta}{2}}\sqrt{\frac{\hat{Pr}(Y)(1-\hat{Pr}(Y))}{n}}$$

which represents the desired upper bound for the error of our estimation. We are able to determine the accuracy of the estimation in relation to n:

$$n = z^2_{1-\frac{\delta}{2}}\frac{\hat{Pr}(Y)(1-\hat{Pr}(Y))}{\epsilon^2} \tag{4.14}$$

Since $\hat{Pr}(Y)(1-\hat{Pr}(Y)) \leq \frac{1}{4}$ then a general bound for n is:

$$n \leq z^2_{1-\frac{\delta}{2}}\frac{1}{4\epsilon^2} \tag{4.15}$$

Equation 4.13 states that our estimate $\hat{Pr}(Y)$ is ϵ close to the real value $Pr(Y)$ with probability less than δ. In this case δ and ϵ are the accuracy and confidence parameters we require. Equation 4.14 gives us a strict bound on the rate of convergence of $\hat{Pr}(Y)$ to $Pr(Y)$. As a direct consequence we notice that values approaching $Pr(Y) = 0.5$ have the slowest convergence. On the other hand Equation 4.12 gives us a more generic with high probability bound for n.

In essence, Equation 4.12 gives us the expected upper bound of the error, which is reaches an accuracy equal to $\epsilon = 0.05$, with $\delta = 0.05$ confidence by

[1] Cumulative distribution function (CDF), or just distribution function, describes the probability that a real-valued random variable X with a given probability distribution will be found to have a value less than or equal to a certain x, where x is one of the values X may attain with a certain probability.

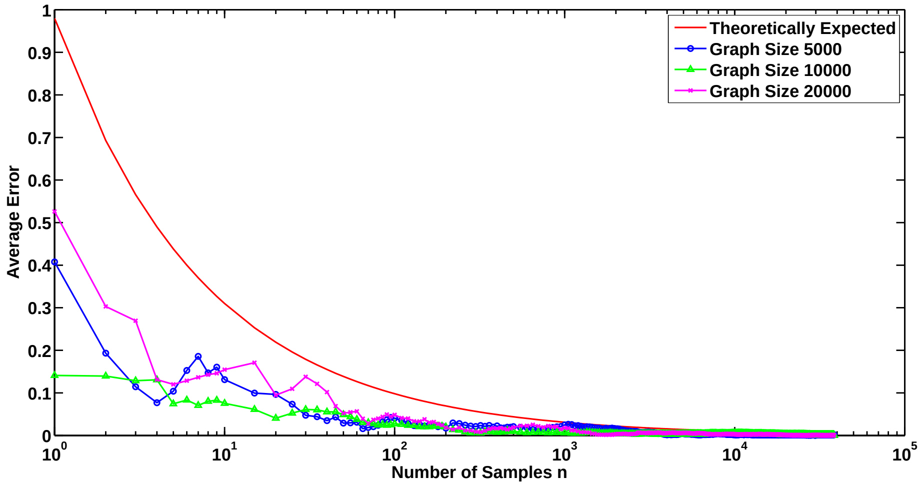


Figure 4.12: Average error over number of samples n

taking:

$$n \leq z_{0.975}^2 \frac{1}{4\epsilon^2} \approx 385 \quad samples$$

independently of the size of the $\mathcal{R}G$. Using Algorithm 4.1 we can estimate the likelihood of a given set of arguments (by dividing their observations by n) and infer a distribution of arguments. Based on these results we can augment our OM by adding to it arguments with probability values above 0.5.

For testing the proposed simulation, we generated Poisson random graphs, with edge probability set to $50/v$ (where v is the size of the graph). The reason for selecting this probability was to ensure there exists a giant connected component and have a graph which is sufficiently dense and coherent to justify the need to use a sampling method to infer the argument likelihoods rather than to directly measure them, but also sufficiently sparse, in order to be able to use it on a computer with limited capabilities. Due to the lack of benchmarks to compare with, we do not know whether a random graph better corresponds to a realistic argument graph, however for the purposes of the Monte-Carlo simulation we believe that the graph structure is irrelevant for the purpose of argument likelihood estimation (assuming that the real argument graphs will not be complete or have a path-like or grid-like structure).

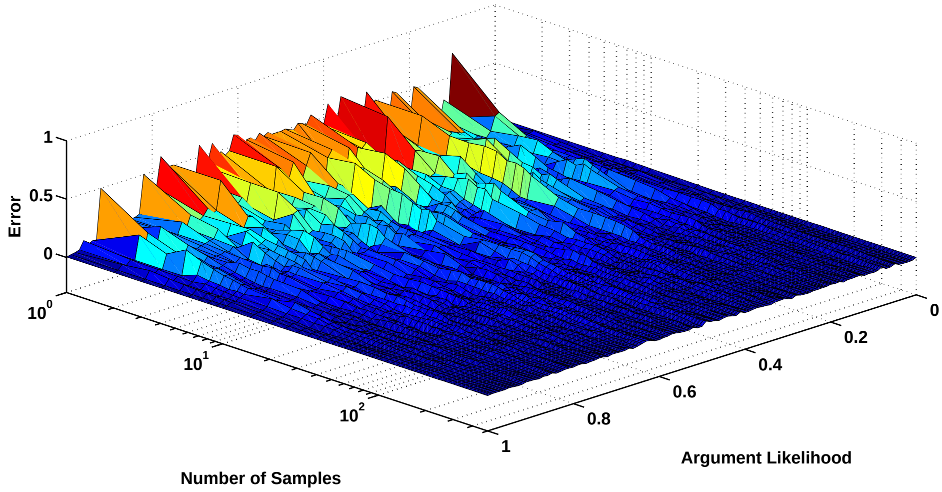


Figure 4.13: Error per argument likelihood over n samples

Performing tests on graphs of various sizes we have obtained the results of the average error which can be seen in Figure 4.12. We can see that the average error is upper bounded by the theoretically expected value. In practice the convergence is much quicker than what the theory suggests ($O(n|R_{\mathcal{A}}|)$ translated to n experiments multiplied with the total number of elements in set $R_{\mathcal{A}}$), resulting in error less than 0.1 after only 15 samples. Additionally in Figure 4.13 we can see the error per argument likelihood $Pr(Y)$. We observe how the error is maximized for likelihoods in the range of $0.4 - 0.6$ which further supports the results of Equation 4.14.

Finally, we make use of a correlation coefficient to measure the correlation between the sampled argument likelihoods with the real likelihoods. In general the use of correlation coefficients is used to determine correlation, or dependence between two random variables. Assuming for example we want to check if there is any correlation between peoples IQ and their GPAs. The correlation coefficient could be an indicator of this correlation, and to obtain it we can first rank each person's IQ and each person's GPAs so as to then test how similar these two rankings are. The idea is that if a person's ranking of his IQ coincides with the ranking of its GPA, e.g. the person with the highest IQ has the highest GPA and so on, then the correlation coefficient would obtain its maximum value (typically

1) indicating that there is a perfect correlation between IQs and GPAs.

In our case we require sampled argument likelihoods to be heavily correlated with the real likelihoods since the first should eventually converge to the second. There are many commonly used approaches to measure the correlation such as:

- The Pearson product-moment correlation coefficient, written as r, which measures the linear relationship between two variables;
- The interclass correlation coefficients used to describe the similarity of units between a specific group of units, and;
- The rank correlation coefficient which measures the linear relationship between the ordering (or ranking) of two variables.

The correlation coefficient we have chosen to use is a rank correlation coefficient, and specifically Spearaman's ρ coefficient. The reason we have chosen to use the ranking correlation coefficient is due to some drawbacks of Pearson's r on very large data[1], underlined by Litvak and van der Hofstad [2012].

Spearman's rank coefficient for fully ranked variables is defined as:

$$\rho = 1 - \frac{6 \sum r_i^2}{n(n^2 - 1)} \tag{4.16}$$

where r_i is the rank difference of the $i - th$ item in the distribution, and n the number of ranked items; in our case, the number of arguments. The Spearman ρ can obtain values between -1 and 1, with -1 meaning that the samples are perfectly anti-correlated and 1 meaning the distributions are perfectly correlated. Values that approach 0 indicate that the distributions are not correlated.

The way we compute the ranking of the items is simply by ordering the arguments by their real likelihood and by their experimentally determined likelihood and comparing the ranks of each argument Y_i in each ordering in order to determine r_i. In Figure 4.14 we can see the correlation of our samples with the real ranks of the nodes in $N_{\mathcal{A}}$. The main point that is apparent in this figure is that the distributions are strongly correlated (over 0.9 correlation) even at small

[1] Particularly, for very large data the second product moment of the coefficient dominates the value, meaning that when used on very large data it will likely result in negative correlation.

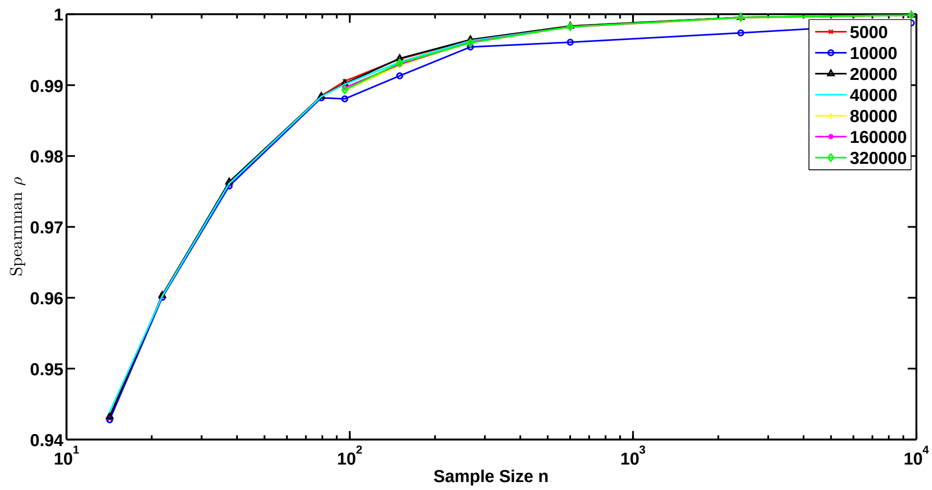


Figure 4.14: Spearman ρ over number of samples n

sample sizes ($k \approx 15$). This means that (empirically) after around 80 samples at most, we should stop sampling since the improvement that would be yielded by additional samples would be of greatly diminished value. Furthermore, the strength of the correlation and the rate of the increase of the correlation is independent of the graph size and does not affect the accuracy of our estimations, which supports the scalability of our approach.

4.5 Augmenting the Opponent Knowledge Base

Given the results of the Monte-Carlo simulation, the next step is to select the neighbouring arguments of the concerned OM with a likelihood higher than a threshold $\theta_S = 0.5$ and augment the concerned OM with the logical constituents of those arguments. Assume for example that an agent Ag_1 augments its opponent model $S_{(1,2)}$ of another agent Ag_2, with the following two arguments:

$$\mathbf{A} : p, s; p, s \Rightarrow q \quad \textit{chosen with a likelihood} \quad Pr(A) = 0.91$$

$$\mathbf{B} : r, t; r, t \rightarrow w \quad \textit{chosen with a likelihood} \quad Pr(B) = 0.84$$

It is then expected that the contents of A and B will become elements of sets $\mathcal{K}_{(1,2)}$ and $\mathcal{R}_{(1,2)}$ according to whether they are premises or rules, while they will also receive a confidence value equal to the likelihood of the argument they compose, as defined in Definition 46(c), i.e.:

$$< p, 0.91 >, < s, 0.91 >, < p, s, \Rightarrow q, 0.91 >$$

$$< r, 0.84 >, < t, 0.84 >, < r, t \Rightarrow w, 0.84 >$$

Nevertheless, this process is not as trivial as it may seem, as even though it may very well be the case that the augmentation concerns arguments new to the model, logical elements of those arguments may already be part of the concerned OM. In this case we face the problem of dealing with conflicting confidence values between those elements. The question raised is whether to either:

- discard the existing confidence value and use the new one, or;
- discard the new confidence value and keep the existing one, or;
- compute an average or a propagated value of the two.

Choosing between one of these three possible ways for resolving these conflicts can be based on various contextual factors, or factors related with the modeller's objectives.

Another issue discussed in this section concerns the instantiation of arguments with logical elements with varying degrees of confidence. Assume for example an argument:

$$\mathbf{C} : f; f \Rightarrow y \quad where \quad < f, 0.87 > \quad and \quad < f \Rightarrow y, 0.63 >$$

then what is the confidence of the produced argument? In other words, what is the confidence with which the modeller assumes an opponent to know a constructed argument? In a similar way to how the conflicting confidence problem can be resolved, numerous approaches can also be employed in this case as well, that are again dependent on the concerned context and on the modeller's goals. In what follows we discuss possible solutions to these problems as well as their advantages and limitations, while we also choose those that better reflect the objectives of our research scope.

4.5.1 Resolving conflicts between confidence values

Since we assume agents to maintain confidence values as a metric for evaluating the likelihood of certain arguments to actually be part of an opponent's knowledge, a naive assumption would be to simply maintain the largest confidence value whenever faced with a conflict. However, this approach seems too general and does not account for the fact that the confidence values result from different information collection methods (ICMs). In other, words one should differentiate between whether a confidence value collected through an augmentation is replaced with a confidence value collected through a third party informer and vice-versa, or with information directly collected through an agent's commitment store.

For example, if certain data are given a confidence value equal to 1 resulting from their direct collection from an agent's commitment store, in the case of conflict their confidence value must not be replaced with a new one. However, in the opposite case it should. In this sense, it is worth defining a precedence hierarchy between the three information collection approaches and the confidence values that result from them. In addition, in the case where a confidence value is the product of a propagated probability acquired through a third party, if it conflicts with a value also acquired from another third party, then, assuming a static trust network, it seems reasonable to discard the lowest of the two—the one we trust the less.

Finally, if the conflict concerns information collected from two distinct augmentations, since the newest of the two will be based on more dialogues, i.e. on more experience, the old one should be discarded. This could also be the case for when a conflict occurs between information collected by third parties—if the concerned trust network is dynamic we should rely on the newest of the two values. Furthermore, again in the case of conflicting confidence values which may result from the same augmentation, i.e. confidence values for the same logical constituent found in discrete arguments included in an OM from the same augmentation, one could choose between the following alternatives: compute an average, discard the lowest or discard the highest of the two.

However the question that remains unanswered is how to deal with conflicts between augmentation confidence values and third party trust related confidence values. Intuitively the difficulty of responding to this question results from the fact that these ICMs appear to be independent, and thus defining a metric able to objectively characterise the two, and thus to choose between them, is at least a complex issue. Even if one assumes that, for example, contradiction between the two could be defined in the sense that it is counter-intuitive to assume that information provided by a single agent could ever be more reliable than one's own general experience suggests, it is perhaps arbitrary to assume that augmentation confidence values should always take precedence over trust related confidence values.

For practical reasons and for the sake of completeness we propose computing an average of the two. In general though, researching the relationship between the implications that follow if an agent strongly trusts a peer, while its experience stands against its trust, appears to be an interesting research problem.

We provide a definition of a function for resolving such conflicts, which also describes how this function can be practically used. We nevertheless note that resolving such conflicts can be done in various ways, which may depend on the nature of the concerned context, or on the perspective and the objectives of the modeller. We thus hold no absolute stance on the appropriateness or on the correctness of this particular approach. In essence, every possible combination of information collection methods (t, t'), out of a total of 2^3 possible combinations, is mapped to a specific operation which reflects the approaches discussed so far. We assume three different information collection methods:

- `dir`: Direct collection
- `tp`: Third party information provider
- `aug`: Augmentation mechanism

and we accordingly define a different operation for all possible combinations which appear in Table 4.4 as described in Definition 56.

Definition 56 (Resolving confidence conflicts) *Let* $S = \{\mathcal{K}, \leq', \mathcal{R}, \leq, \mathcal{G}\}$ *and* $< x, c, t >$ *be a tuple, where* $x \in \{\mathcal{K}, \mathcal{R}\}$*,* c *is* x*'s corresponding confidence value, and* $t \in ICM$ *where:*

$$ICM = \{\texttt{dir}, \texttt{tp}, \texttt{aug}\}$$

Assume then that $< x, c', t' >$ *is an element to be incorporated into* S *such that* $c \neq c'$*, then* f_c *is a function where:*

$$f_c : t \times t' \rightarrow [0, 1]$$

through which we decide on the final confidence value c *associated with* x *as follows:*

$$f_c(c, c', t, t') = \begin{cases} c & \textit{if } t = \texttt{dir} & (a) \\ c' & \textit{if } t' = \texttt{dir} & (b) \\ max\{c, c'\} & \textit{if } t = \texttt{tp} \textit{ and } t' = \texttt{tp} & (c) \\ c' & \textit{if } t = \texttt{aug} \textit{ and } t' = \texttt{aug} & (d) \\ \frac{c+c'}{2} & \textit{if } t = \texttt{tp} \textit{ and } t' = \texttt{aug} \textit{ and vice versa} & (e) \end{cases}$$

Notice that in the case where both t as well as t' are equal to `dir` it is also the case that both c and c' will be equal to 1 and thus to each other, in which case there is no conflict.

4.5.2 Computing an argument's confidence value

In a similar sense to the approaches proposed for resolving possible conflicts between the conflicting confidence values, in the case of computing an argument's confidence value, one may rely on a variety of ways. These may concern:

- computing the propagated confidence value resulting from the multiplication of the constituents of a particular argument;
- computing an average of confidence value of the constituents of an argument; or

t	t'	**Result**
`dir`	`tp`	Keep c
`dir`	`aug`	Keep c
`tp`	`tp`	$max\{c, c'\}$
`tp`	`dir`	Keep c'
`tp`	`aug`	$\frac{c+c'}{2}$
`aug`	`aug`	Keep c'
`aug`	`dir`	Keep c'
`aug`	`tp`	$\frac{c+c'}{2}$

Table 4.4: The possible combinations of the three information collection methods and how they are resolved in the case of conflict, where: `dir`: Direct collection; `tp`: Third party information provider, and; `aug`: Augmentation mechanism.

- assigning a confidence value equal to either the largest or the smallest confidence value of the argument's constituents.

However, another issue needs to be raised at this point; one which is more related to the employed logic system and the nature as well as the semantics of the encoded logical information used in that system.

In general, in the *ASPIC*$^+$ framework the construction of an argument is partly based on the use of deductive principles that allow for deductive inferences of conclusions from premises. Examples of such principles are modus-ponens or modus-tollens ($p \to q, p \vdash q$ respectively $p \to q, \neg q \vdash \neg p$) which generally represent inference rules that exist independent of the logical language. The set of strict inference rules ($\mathcal{R}_s$) encodes such classical deductive inferences. At the same time these rules can be expressed as elements of the language, i.e. be represented as forms of axiomatic premises resulting from material implication, (e.g. $\alpha, \alpha \sqsupset \beta \to \beta$) whose validity is indisputable ($\mathcal{K}_n$). Thus, arguments constructed based on deductive or strict inference rules can be based on premises that are equally valid to strict or deductive rules in the sense which are also unquestionable. In this case, the validity of such arguments cannot be compromised, while since these arguments are based on axiomatic beliefs as well as on deduction rules that are independent of the employed logic (though expressed as elements of it) these arguments should be perceived as commonly known arguments and be given a confidence value equal to 1. In other words, it makes sense to assume that all agents in a multi-agent environment partly share such 'absolute' information (we refer the reader to Definition 18, Section 3.1).

$ASPIC^+$ additionally allows for the construction of arguments based on defeasible inference rules ($\mathcal{R}_d$), which can be attacked on both their conclusions and premises, as well as on their inference steps. The distinction between premises that are axiomatic and that cannot be attacked ($\mathcal{K}_n$) from other forms of premises with questionable validity and susceptible to attacks ($\mathcal{K}_p \cup \mathcal{K}_a$), as well as between strict and defeasible rules ($\mathcal{R}_s$ and $\mathcal{R}_d$) is very useful. One may rely on this distinction for assigning confidence values to constructed arguments, which, based on their logical constituents, appear to be axiomatic and which should thus be assumed to be part of a 'universally' shared knowledge, and thus to distinguish them from other arguments whose validity is questionable and for which a confidence value should be calculated and assigned to them.

However, the $ASPIC^+$ framework also allows for domain specific representation of strict inference rules and and axiom premises; something which adds to the framework's flexibility. To provide a simple example, assuming that:

$$\texttt{bachelor} \rightarrow \neg\texttt{married}$$

is a strict inference rule or a defeasible one is a debatable issue. In either case, one should be flexible enough to decide which ever better reflects one's perception. We generally assume that agents' OMs allow for any kind of premises and rules. Therefore the association of any kind of a logical constituent with a confidence value makes more sense, as the concerned constituent can no longer be perceived as being part of a shared universal knowledge base, but is rather one's own personal belief.

In general, each of the proposed approaches presented in the beginning of this section, relies on different perspectives and accounts for different inherent properties that should characterise an argument's confidence value. For example the first case follows from basic probability theory laws which state that the probability of an event taking place is equal to the propagated probability of the events it depends on. In this sense it is only reasonable that the confidence value of an argument:

$$C : f; f \Rightarrow y \ where \ < f, 0.87 > \ and \ < f \Rightarrow y, 0.63 >$$

is equal to $0.87 \cdot 0.63 = 0.5481$. In addition the employment of this approach results in defining a probability distribution over the possible combinations of the constituents of an argument, with respect to the possibility of them being actually part of the opponent's knowledge base or not, i.e.:

$$\begin{aligned} Pr(f) \cdot Pr(f \Rightarrow y) &= 0.5481 \\ Pr(f) \cdot (1 - Pr(f \Rightarrow y)) &= 0.3219 \\ (1 - Pr(f)) \cdot Pr(f \Rightarrow y) &= 0.0819 \\ (1 - Pr(f)) \cdot (1 - Pr(f \Rightarrow y)) &= 0.0481 \end{aligned}$$

where:

$$\sum_{x \in A} Pr(x) = 1$$

where A is an argument instantiated from a knowledge base S and x is a logical constituent of A.

A similar approach to this has been proposed by Hunter [2013]. In his work Hunter looks into the logical constituents of arguments as well for defining a probability value for characterising the argument's uncertainty, though in his case the objective is to decide the extend to which an argument is true in general, in addition to the fact that he assumes arguments constructed based on classical logic inferences, i.e. only assumes strict inference rules in the language. In our case we care to find out the certainty level with which one should believe that its opponents are aware of certain information.

Though propagation appears to be a reasonable approach for deciding on the confidence of a constructed argument, there is a problem with its employment in our case. Namely, that it produces an antisymmetry between the confidence value of an argument whose confidence value is also assigned to its constituents and the confidence value of the same argument after its reconstructed from those constituents. Assume for example that an argument's likelihood to be known to a certain opponent satisfies an augmentation threshold set at 0.5. Let that argument be A, then this likelihood value will be directly transferred to its constituents:

$$< A : p; p \Rightarrow q, c_A = 0.5 > \ \ where \ \ < p, c_p = 0.5 > \ \ and \ \ < p \Rightarrow q, c_{p \Rightarrow q} = 0.5 >$$

which will then be added to that opponent's OM. Reconstruction of that same argument from its constituents, and computation of its confidence value will result in an overall argument confidence value of 0.25, which will be half of the original value which one would expect it to have.

One could imply that this suggests that assigning the constituents of an argument with the likelihood value of an argument is also false. If the construction of an argument is based on whether its constituents are both part of a knowledge base, then knowledge of each of those constituents should be perceived as a random events with discrete probabilities, while the constructed argument should be their propagated result. In other words, if one assumes for example that an argument's likelihood of being known to a certain opponent is equal to 0.5 then that number should be the product of the propagated likelihood values of its constituents, e.g.

$$< A : p; p \Rightarrow q, c_A \approx 0.5 > \ \ where \ \ < p, c_p = 0.91 > \ \ and \ \ < p \Rightarrow q, c_{p \Rightarrow q} = 0.55 >$$

However there are infinite combinations of numbers whose product will be equal to 0.5. In addition, our augmentation approach relies on an abstract perception of arguments and it thus disregards likelihood information related directly to the constituents of arguments. In other words deducing the likelihood of the constituents of arguments rather than of the arguments constructed by them is a different issue, is not accounted in our approach. We elaborate more on this issue in Chapter 5.

One could try computing the average of the confidence values of the contents of an argument, as it also appeals to intuition. It is generally a common approach to assume that the volatility of a possible event can only belong in the range between the highest or the lowest probability of the events it depends on. It can thus be expressed through the average value of those probabilities. However the same asymmetry issues arises here as well.

We therefore resort to, the last approach, which resembles the *weakest link principle* presented by Prakken [2010] for deciding on whether a certain argument defeats another, in the case where the smallest of the confidence values of the constituents of an argument is selected to represent the argument's confidence value. Intuition here lies in the fact that if an event depends on a number of other events in order to happen, then in essence its probability of happening is at most equal to the probability of the event with the lowest probability. In contrast to the two approaches previously mentioned, in this case the confidence value of the argument from which the values of the constituents are inherited and the confidence value of the same argument reconstructed from those constituents will be the same. We note that the symmetry between the original argument's confidence value and the confidence value of that argument after it has been reconstructed from those constituents should only hold in the case where no confidence value conflicts occur between the newly imported constituents and constituents which are already part of an OM, as in such case resolution of conflicts will result in altering those values, as explained in Section 4.5.1.

We thus opt to rely on the latter approach for defining the confidence value of an argument composed of logical constituents with varying degrees of confidence. Nevertheless, we provide two general formulæ for defining an argument's confidence value as follows:

Definition 57 (Argument confidence value) *Let A be an argument and $\Lambda = \{x_1, x_2, \ldots, x_i, \ldots, x_n\}$ be a set containing A's logical constituents where $\forall x$ holds that $x \in \{\mathcal{K}, \mathcal{R}\}$ while there exists a tuple $< x_i, c_i >$ where c_i is the confidence value of x_i, for $i = 1, 2, \ldots, n$. Then the confidence value c_A of the argument composed from the elements of Λ is computed based on the following general formula:*

$$c_A = c_1 \circledast c_2 \circledast \ldots \circledast c_n$$

where $\circledast$ is a user-defined operator.

Definition 58 (Weakest Link Confidence) *Let A be an argument and $\Lambda = \{x_1, x_2, \ldots, x_i, \ldots, x_n\}$ be a set containing A's logical constituents where $\forall x$ holds that $x \in \{\mathcal{K}, \mathcal{R}\}$ while there exists a tuple $< x_i, c_i >$ where c_i is the confidence value of x_i, for $i = 1, 2, \ldots, n$. Then the confidence value c_A of the argument composed from the elements of Λ is computed based on the following general formula:*

$$c_A = min\{c_1, c_2, \ldots, c_n\}$$

Finally in relation to the user-defined operator used in Definition 57 we note that, though it is not necessary, the operator should provide a result within the closed range of $[0 - 1]$ in order to be used as a probability value for decision making purposes. However, in the case where this requirement is not satisfied, the user may normalise the c_A values to represent probabilities at a later stage.

4.6 Conclusions & Contributions

In this chapter we have provided a general methodology for updating and augmenting an OM, based on an agent's experience obtained through dialogues. This methodology is based on two mechanisms respectively responsible for updating and augmenting an OM. In relation to the latter, we provided a method for building a graph between related arguments asserted by a modeller's opponents, referred to a $\mathcal{R}G$, and proposed an augmentation mechanism, enabling an agent to augment its current beliefs about its opponents beliefs by including additional information (arguments), that is of high likelihood to be related to what the opponent is currently assumed to know. Thus, we enabled an agent to also account in its strategising for the possibility that additional information may also be known to its opponents, while indirectly accounting for the structure of the dialogues from which information is collected—i.e. accounting for how and when certain opponent arguments seem to follow after certain others. We thus allowed a modeller to utilise her dialogue history in a multifaceted way. Lastly, we defined and analysed a Monte-Carlo simulation which enabled us to infer the likelihood of those additional arguments in a tractable and efficient way.

We are aware that more investigation is needed with respect to relying on alternative contextual factors for quantifying the likelihood between elements in a $\mathcal{R}G$, such as the level of a participant's membership in a group, and this is something we intend to investigate in the future. In addition, it is worth to also investigate whether the links between arguments in a $\mathcal{R}G$ can be attached based on different approaches, other than relying on the notion of support. For example, it is worth to investigate whether arguments should be related—and therefore attached to each other—given that they allegedly attack the same arguments, or if they are partly composed of similar constituents, i.e. it is more likely for someone to be aware of an argument $B : p, p \Rightarrow q$ given that he knows $A : p, p \Rightarrow s$ as they are both aware of premise p. Furthermore, it is also essential that we investigate the added value of modelling a $\mathcal{R}G$ for a $\theta_t > 1$, as increasing θ_t's value may significantly increase the complexity of the proposed solution. In a similar sense, it seems also interesting to investigate whether an agent should attempt to augment its knowledge with arguments that are at more than a one hop distance from a concerned OM, as the proposed mechanism is only concerned with neighbouring nodes of the model. Of course this implies the need for the development of a different augmentation approach. All these issues are both discussed and dealt with in Chapter 5.

Another issue is concerned with the evaluation of the effectiveness of our approach which requires that we take into account actual argumentation frameworks and not just abstract arguments. Unfortunately, there is a general lack of benchmark mechanisms in the field of argumentation and this is an interesting issue which we intend to tackle in the future. A methodology towards evaluating our approach is discussed in Chapter 7.

It is worth noting that the proposed approach can be easily generalised and applied in different contexts. All that is necessary is the development of a methodology for associating (linking) information (locutions) based on contextually inherent properties. For example sharing access to certain information could concern a trial case where one would expect that the participating lawyers are equally aware of facts concerned with the trial. In this case, assuming that the modelling subject is a lawyer participating in a trial, then we may augment his model by adding to it information we believe other lawyers, who also participate in the trial, know. Another example could concern information related to a particular topic. Let us assume that someone is aware of the fact that "earth is a sphere" and that "it is part of the solar system", this gives reasonable grounds for also assuming that, with a certain likelihood, it is possible that that someone also knows that "earth is the third planet orbiting around the sun", given of course that we can first categorise a modeller's collected information into topic groups. Finally, one could even rely on 'reasonable assumptions' for inferring a relationship between information. Assume for example that an academic is aware of a certain paper, it is then reasonable to assume that she is likely to also be aware of papers cited in it, while one would expect that this likelihood would increase if she is the author of the paper. As far as our work is concerned this inherent linking property is expressed through the notion of support. More approaches to associating potentially related information based on whether it is characterised with other forms of logical interrelatedness is discussed in the following chapter.

Chapter 5

Extending Opponent Modelling

This chapter focusses on extending the opponent modelling methods presented in Chapter 4, which rely on the construction of relationships graphs ($\mathcal{R}G$s). Such graphs can be used for augmenting one's knowledge of the possible knowledge of its opponents, based on the relationships between the opponent arguments (OAs) in a $\mathcal{R}G$.

Particularly, we extend the weight assignation mechanism responsible for assigning weight values on the arcs which link OAs, that allows for a more concrete modelling approach to quantifying these values, accounting for additional aspects that were ignored in Chapter 4 for the sake of simplicity.

We also discuss different ways for building $\mathcal{R}G$s, i.e. ways for relating OAs based on inherent relationships they share. In Chapter 4, the inherent relationship used was the notion of *support*. We introduce a new relationship notion which allows the formation of links between arguments with *common attack targets*, thus utilising another inherent linking property between arguments. We also illustrate how both notions can be combined for the construction of a more dense[1] $\mathcal{R}G$.

5.1 Introduction

Out of the three modelling mechanisms presented in Chapter 4, direct collection, through the opponent's commitment store (`dir`); third party provided informa-

[1]Here "dense" is not use in its standard graph theoretic sense but rather as a placeholder for "has more links".

tion (`tp`), and; augmentation (`aug`), we mostly focussed on the last one. The augmentation mechanism is used to expand an agent's current model of its opponent, by adding to it information that has a high likelihood of being related with information already contained in it. As this is a complex mechanism which is required for numerous new notions to be introduced and explained, we purposely opted to present a simple version of the proposed methodology which, as it will become evident in this chapter, disregards many important considerations which if were taken into account could have a considerable impact on the effectiveness[1] of the proposed modelling methodology, i.e on how valid our modelling approach can be.

As discussed in Chapter 4, for linking arguments in a $\mathcal{R}G$ we rely on their interrelatedness. In other words, we rely on inherent relationships that arguments may share for associating them with some relationship likelihood. We can then rely on these relationships in order to anticipate possible arguments that could follow in a dialogue game, if they are related with arguments already in the game, given that this has repeatedly been recorded in past dialogues with different opponents. Such inherent relationship is the arguments' ability to reinstate one another, which we expressed through the notion of support. We differentiate between two cases of support. In the first case, assuming an argument C that attacks an argument B which in turn attacks another argument A, then we assume that C directly supports A as it *reinstates* its acceptability by attacking B. The second case concerns a weaker version of support, referred to as *indirect support*, in which case another argument, D, would be introduced in the same dispute as A by the same participant as A but not for attacking A's direct attacker B, i.e. intermediate arguments (e.g. C) introduced by that participant appear between A and D in a dispute.

Let us also recall the weight assignation formula presented in Definition 52 responsible for the quantification of the support relationships between arguments in a $\mathcal{R}G$. According to this formula, assuming two related—linked—arguments A, B in a $\mathcal{R}G$, a weight value within the closed range of $[0, 1]$ is produced by counting the number of times that B follows after A in discrete dialogues and

[1]With effectiveness here we refer to the mechanism's ability to better reflect one's opponents' actual knowledge, or to better anticipate the likelihood of something to follow in a dialogue.

dividing that number with the total appearances of A in dialogues. One should agree that this is a rather abstract approach for quantifying the relationship likelihood between the connected arguments in a $\mathcal{R}G$. For example, one could adjust the weighing value so as to account for the θ-distance between a linked pair of OAs in a way analogous to how strongly the following argument supports the preceding argument. In other words, the weighing formula should be able to discriminate between cases of linked pairs of OAs where one directly supports the other (reinstatement) and pairs where one indirectly supports the other. Of course, in Chapter 4 we consider only $\mathcal{R}G$s built based on a $\theta_t = 1$. As discussed, increasing the value of the θ_t results in increasing the connectivity between the arguments in the $\mathcal{R}G$, and therefore the complexity is also increased. So, unless there is some added value in setting a θ_t with values greater than 1, this should be avoided. This is one of the issues we investigate in this chapter.

Furthermore, recall the connectivity condition presented in Definition 51. The condition dictates that arguments should be connected only if they appear in the same line of dispute. This is a condition imposed in order to reflect the support relationship between OAs. However, support is not the only inherent relationship between OAs. Another such inherent relationship could be found in how often a set of arguments appears to attack the same target—argument—in an agent's history. Specifically, we could assume that arguments with *common attack targets* could be associated with a likelihood deriving from how often they appear to attack their common target, in a history of dialogues. After all, it seems reasonable from a modeller's perspective to hold the belief that if an opponent is aware of a single argument able to counter one of its own, then that opponent may also know more arguments that can do the same.

The rest of the chapter is structured as follows: in Section 5.2 we discuss the added value of building $\mathcal{R}G$s with $\theta_t > 1$, we propose a different weighing mechanism which accounts for additional modelling aspects than the original mechanism presented in Chapter 4, and we compare the two mechanisms; in Section 5.3 we propose a different modelling approach for building $\mathcal{R}G$s, assuming relationships between arguments with *common attack targets* and finally; in Section 5.4 we summarise our contributions.

5.2 Support–$\mathcal{R}G$s with $\theta_t > 1$

In Section 4.3.1 we discuss and propose a modelling approach for building a $\mathcal{R}G$ which relies on the notion of support. Particularly, arguments are assumed to be related if they are found in the same dispute lines of a dialogue and at a θ-distance equal to θ_t from each other. This distance (θ_t) was purposely set to 1, in order to easily introduce the proposed modelling approach, capturing support in its strongest form—that of reinstatement. However, a question worth investigating is whether there is any added value in assuming an indirect support relationship between arguments—one that exceeds the θ_t threshold of 1.

Increasing the value of θ_t for modelling a $\mathcal{R}G$ relies on the assumption that opponent arguments (OAs) found in the same line of dispute are related, since one supports the other regardless of the distance between them. Though this seems like a valid assumption there are numerous considerations that one is called to account for prior to applying it in modelling a $\mathcal{R}G$.

Firstly, increasing θ_t will result in building a more dense $\mathcal{R}G$ which means that the complexity of computing the argument likelihoods of the augmentation process will also increase. In other words, increasing the density of the $\mathcal{R}G$ graph may make the augmentation process intractable, while at the same time from a human point of view not bounding θ_t ($\theta_t \approx \infty$) or assigning a large value to it, raises a cognitive resources issue, due to the large volume of information that needs to be stored. In addition to these though, and more importantly, it is semantically questionable whether an argument Y, that follows at a large distance after another argument X in the same dispute, actually does support X. Therefore, it is at least necessary that with the increase of θ_t corresponding weighing mechanisms are developed in order to appropriately quantify the relationship likelihood between the supported arguments, in such a way so as to account for the distance between them.

For example, let us re-examine the modelling example first presented in Chapter 4, which appears again in Figure 5.1. Figure 5.1c illustrates a $\theta_t = 2$ modelling approach where, in contrast to Figure 5.1b which assumes a $\theta_t = 1$, shows a link in the form of a direct arc between arguments B and F. Let us now consider the weighing mechanism according to which the weight values w_{BF} and w_{BD} are

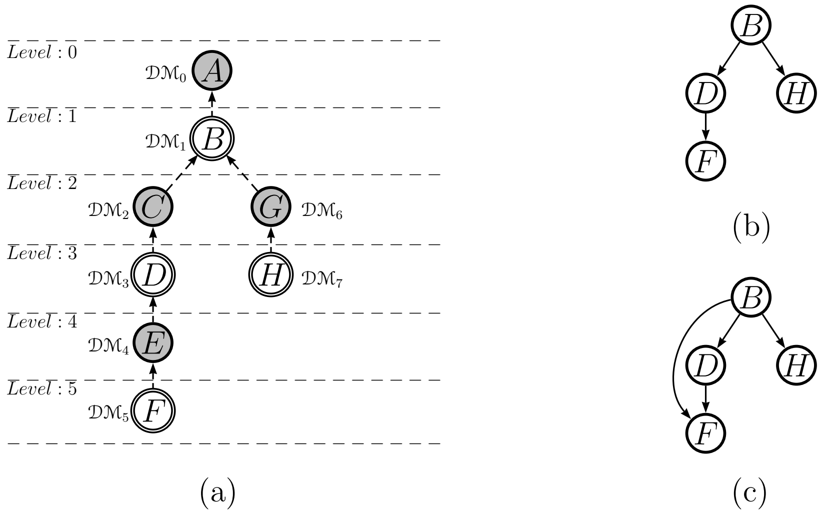


Figure 5.1: a) A dialogue tree $\mathcal{T}$ where the grey and the white nodes concern Ag_1's respectively Ag_2's moves, b) A 1-hop $\mathcal{R}G$ modelling approach c) A 2-hop $\mathcal{R}G$ modelling approach.

decided. We remind the reader that weights on the arcs have values in the closed range of $[0, 1]$ and are assigned as in Definition 52, in which the weight is the number of times that the *end*-argument (in this case F and D) follows after the *start*-argument (in this case B), divided by the total appearances of the *start*-argument. If we assume that argument F was recorded to only follow in dialogues after argument D then its dependence on D should somehow be reflected in the corresponding weights on the edges that link arguments B, D and F. However, according to the proposed weighing approach the weighing values of the pairs B, D and B, F will be exactly the same, as they will both have the same numerator (the number of times that F follows after B is equal to the number of times that D follows after B), and the same denominator (the total number of times that B appears).

Practical issues aside, intuition suggests that there is little to learn by increasing the θ_t threshold when modelling an $\mathcal{R}G$ while it may produce counter-intuitive results. For example, since the indirect support of argument B by F as a possibility is dependent on the possible awareness of D, it makes little sense to augment an OM with F but without D. Nevertheless, this becomes possible in the case where $\theta_t = 2$ in contrast to when $\theta_t = 1$ due to the existence of r_{BF}, even if F only appears to follow after D in dialogues. Though appropriate weighing quan-

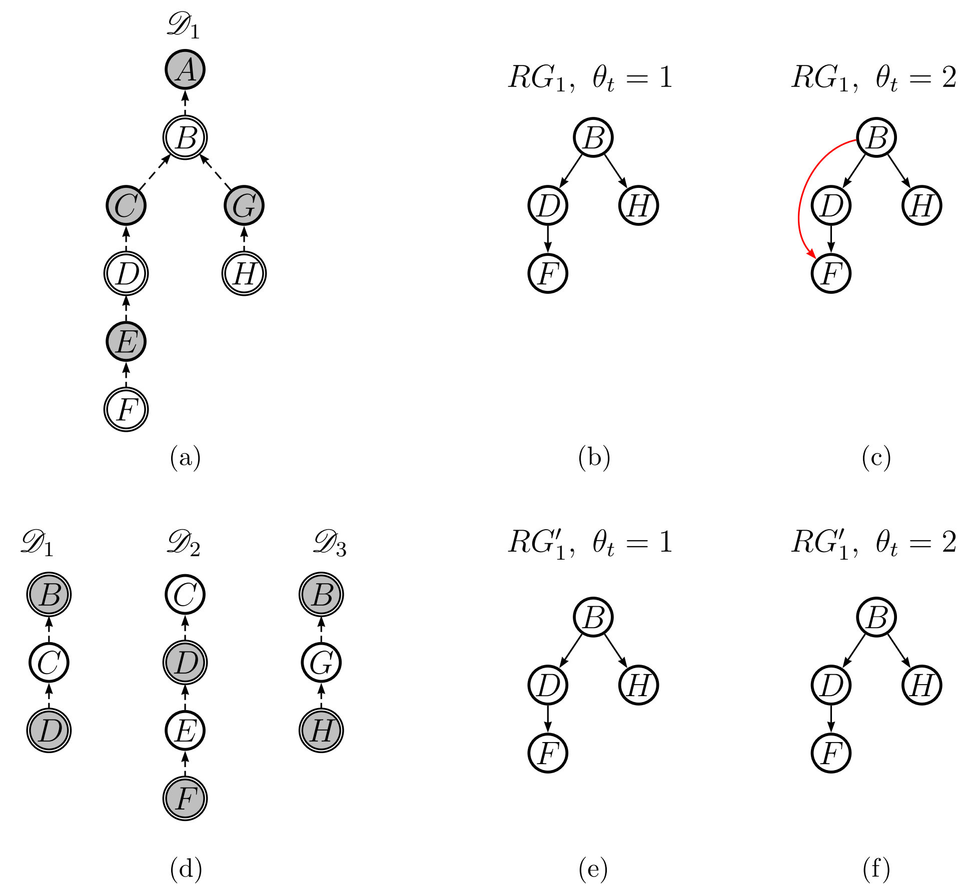


Figure 5.2: a) A dialogue between an agent Ag_1 (grey) and an agent Ag_2, b) The induced $\mathcal{R}G_1$ for $\theta_t = 1$, c) The induced $\mathcal{R}G_1$ for $\theta_t = 2$, d) Three dialogues between Ag_1 and Ag_2, e) The induced $\mathcal{R}G_1'$ for $\theta_t = 1$, and f) The induced $\mathcal{R}G_1'$ for $\theta_t = 2$.

tification can diminish this possibility, it cannot exclude it completely. Thus, increasing θ_t when building an $\mathcal{R}G$ seems both troublesome and not worth the effort to do so, unless there is more to learn.

In fact, there is one more important thing we can learn by increasing θ_t which we elaborate in the following example.

Example 8 *Let us assume that two agents Ag_1 and Ag_2 engage for the first time in a single dialogue $\mathscr{D}_1$ (Figure 5.2a), based on which an $\mathcal{R}G_1$ is induced for θ_t=1 (Figure 5.2b) and another is induced for a θ_t=2 (Figure 5.2c) . Let us then assume another case where the same agents engage, again for the first time, in a number of dialogues instead of just one, $\mathscr{D}_1, \mathscr{D}_2$ and $\mathscr{D}_3$ (Figure 5.2d), based* *on which another $\mathcal{R}G'_1$ for θ_t=1 is induced (Figure 5.2e) as well as for a θ_t=2 (Figure 5.2f).*

In both cases, for a θ_t=1 the structure of the induced $\mathcal{R}G$, is the same, i.e. $\mathcal{R}G_1 = \mathcal{R}G'_1$. However, if we were to induce these $\mathcal{R}G$s for a θ_t=2 then, in contrast to $\mathcal{R}G'_1$ of Figure 5.2f which would remain the same, $\mathcal{R}G_1$, depicted in Figure 5.2c, would have an additional arc (the red arrow) connecting arguments B and F.

Based on this example, it is easy to see that if we were given just an $\mathcal{R}G$ of the form presented in Figures 5.2b & 5.2e , then assuming that this $\mathcal{R}G$ was modelled for a θ_t=1, we *would not* be able to distinguish between whether B and F appeared in the same dialogue or not. In other words, distinguishing between whether the concerned $\mathcal{R}G$ is a result of a single dialogue such as the one illustrated in Figure 5.2a, or of numerous distinct dialogues such as those illustrated in Figure 5.2d would not be possible, since for a θ_t=1 the induced $\mathcal{R}G$s would be identical in both cases. In contrast, if we build the $\mathcal{R}G$ for a θ_t=2, an additional arc will be added between arguments B and F in the case where B and F have actually appeared in the same dialogue (the red arc in Figure 5.2c).

Semantically, increasing the value of θ_t is important if we want to be more precise in relation to better capturing the relationship likelihood between two arguments. However, the added value of this modelling approach is not just that it allows us to distinguish between whether two arguments have or have not appeared in the same dialogue, by the respective existence or non-existence of an arc between them, but that it captures an additional modelling objective which derives from our modelling hypothesis. Namely, that in contrast with arguments that do not appear in the same dialogue, those that do should be characterised by a larger relationship likelihood. In other words, the likelihood of someone knowing F given that she knows B should increase if they have both appeared in the same dialogue.

After all our objective is to anticipate what arguments might follow in a dialogue after certain others, and thus if two arguments do appear in the same dialogue it is in our best interest to record/represent this in our modelling. In this case, this is captured through the addition of an arc which directly links B and F, since in this way the probability of getting from B to F, through a *random walk*, can only increase, as there are now more ways to get to F from B. Recall however,

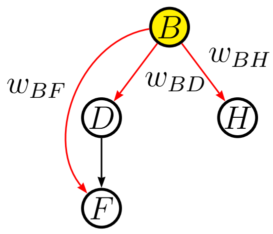


Figure 5.3: The possible expansion options of a sub-graph that contains only argument B.

that our augmentation approach attempts only single-hop graph expansions. In other words, assuming the $\mathcal{R}G$ that appears in Figure 5.3 where the OM to be expanded is one that only contains argument B, the only way to get to F is through r_{BF}, and the probability to get to F in this case is exactly equal to the edge's corresponding weight, i.e. $P_{BF} = w_{BF}$. So stating that we have increased the number of ways we can get to an argument that indirectly supports another, does not really hold. Nevertheless, we have in fact increased the probability of getting to such an argument (in the case of our example, to get from B to F) from 0—since there was no direct link between arguments that indirectly support one another before—to the weight value (w_{BF}) of the arc that links that argument to the one it indirectly supports (B). What we intend to additionally account for now, is the fact that such indirect relationships should have lower probabilities from the direct relationships they extend. In other words, that the weight w_{BF} is lower than w_{BD}, even if F always follows after D in dialogues.

The intuition behind this modelling approach lies in the fact that assuming that the two weights are equal ($w_{BF} = w_{BD}$), would disregard that F following B is contingent on the intermediate D being moved. In other words, if F always follows after D in dialogues but never on its own after B, then assuming a weight value $w_{BF} = w_{BD}$ would completely ignore this fact. So far we rely on a naive approach for providing a weight value w_{XY} for a corresponding arc r_{XY}, simply counting the number of times that an argument X followed by Y appeared in dialogues, and dividing them with the total number of times that argument X appeared in dialogues. This approach disregards that argument Y might appear in various distances from X. If we were to rely on this approach then, in the case of our example, w_{BF} and w_{BD} would be exactly equal if F always follows after D. We therefore need to also take into account the distance between the concerned arguments. This can be done through monitoring both the number of times that a certain argument follows after another in a dispute as well as the distance they have from each other, and using that distance for lowering the relationship likelihood between the associated arguments. To do so we rely on the following definition:

Definition 59 (Relationship Instances) *Given an Ag and its $\mathcal{RG} = \{\mathcal{A}^{\mathcal{H}}, R\}$, its history of dialogues $\mathcal{H}$, and two arguments A, B elements of $\mathcal{A}^{\mathcal{H}}$ then let:*

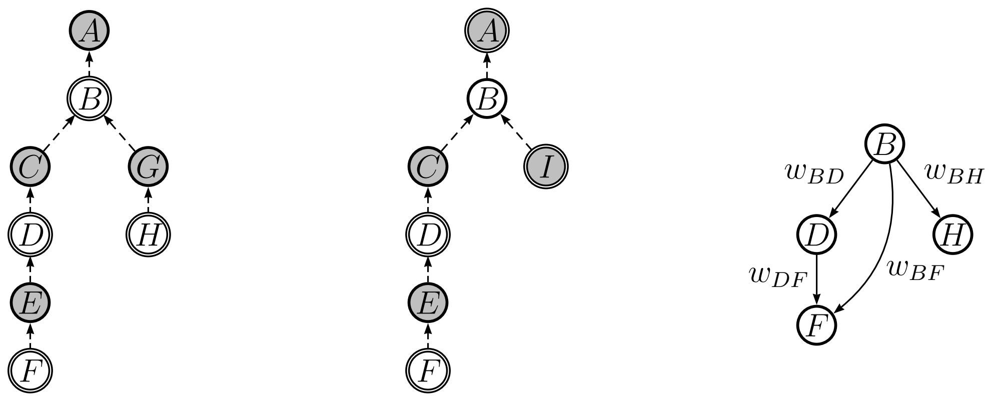
$$\texttt{Instances}(\mathcal{H}, A, B) = \mathcal{I}_{AB} = \{i_1, i_2, i_3, \ldots, i_\mu\}$$

be a function that returns a set $\mathcal{I}_{AB}$ representing all the instances i of argument A followed by B in the same disputes, satisfying Definition 51 such that $r_{AB} \in R$, and where each $i_k \in \mathcal{I}_{AB}$ is equal to the θ-distance between A and B in that instance[1]*.*

Essentially, through this definition, we are able to not only count the instances of two arguments in an agent's history of interactions, but also the corresponding distances they have. Note that it is always the case that $|\mathcal{I}_{A*}| \geqslant |\mathcal{I}_{AB}|$, where $*$ is substituted with any argument in $\mathcal{A}^{\mathcal{H}}$ including the `null` argument, as the instances of argument A will be at least equal to the instances of A followed by B. Assume for example the two dialogues of Figure 5.4 between Ag_1 (the modeller) and another two agents, Ag_2 and Ag_3. From these two dialogues the modeller will induce the $\mathcal{RG}$ which appears in Figure 5.4c for a $\theta_t = 2$. The corresponding instances of the pairs joined by the 4 edges in the induced $\mathcal{RG}$, r_{BD}, r_{DF}, r_{BF}, and r_{BH} are:

$$\mathcal{I}_{BD} = \{1, 1\} \qquad \mathcal{I}_{BF} = \{2, 2\}$$
$$\mathcal{I}_{DF} = \{1, 1\} \qquad \mathcal{I}_{BH} = \{1\}$$

[1]Note that in a single dialogue there may exist more than one instances of arguments A and B with varying distances.



(a) (b) (c)

Figure 5.4: a) $\mathscr{D}_1$ betweeen Ag_1 and Ag_2, b) $\mathscr{D}_2$ betweeen Ag_1 and Ag_3, c) the induce $\mathcal{R}G$

while the total instances of arguments B and D are:

$$\mathfrak{I}_{B*} = \{1, 1, 1, 1\} \qquad\qquad \mathfrak{I}_{D*} = \{1, 1\}$$

Given the instances of a pair of connected arguments, we can define a mechanism for assigning a weight value on their corresponding arc in a $\mathcal{R}G$, as follows:

Definition 60 (Conductivity-based Weight Assignation) *Given an agent Ag and its $\mathcal{R}G = \{\mathcal{A}^{\mathcal{H}}, R\}$, its history of dialogues $\mathcal{H}$, and two arguments A, B elements of $\mathcal{A}^{\mathcal{H}}$, such that there exists an $r_{AB} \in R$, then:*

$$w_{AB} = \frac{\displaystyle\sum_{i_k \in \mathfrak{I}_{AB}} \left(\frac{1}{i_k}\right)}{|\mathfrak{I}_{A*}|} \tag{5.1}$$

.

Note that since every instance $i_k \in \mathfrak{I}_{AB}$ and $i_k \in \mathbb{N}$, it must hold that $0 < \frac{1}{i_k} \leq 1$. Therefore, since $|\mathfrak{I}_{A*}| \geqslant |\mathfrak{I}_{AB}|$ it must also hold that:

$$\sum_{i_k \in \mathfrak{I}_{AB}} \left(\frac{1}{i_k}\right) \leq |\mathfrak{I}_{A*}|$$

Essentially, by dividing $\sum_{i_k \in \mathfrak{I}_{AB}} \left(\frac{1}{i_k}\right)$ with $|\mathfrak{I}_{A*}|$, which represents the number of all the instances of argument A, we normalise the sum of the distances of all the instances of A and B in $\mathcal{H}$ thus computing a probability value. In relation to the example of Figure 5.4 the respective weight values for edges r_{BD}, r_{DF}, r_{BF}, and r_{BH} would be:

$$w_{BD} = \frac{\sum\limits_{i_k \in \mathfrak{I}_{BD}} \left(\dfrac{1}{i_k}\right)}{|\mathfrak{I}_{B*}|} = \frac{\frac{1}{1}+\frac{1}{1}}{4} = \frac{2}{4} \qquad w_{BF} = \frac{\sum\limits_{i_k \in \mathfrak{I}_{BF}} \left(\dfrac{1}{i_k}\right)}{|\mathfrak{I}_{B*}|} = \frac{\frac{1}{2}+\frac{1}{2}}{4} = \frac{1}{4}$$

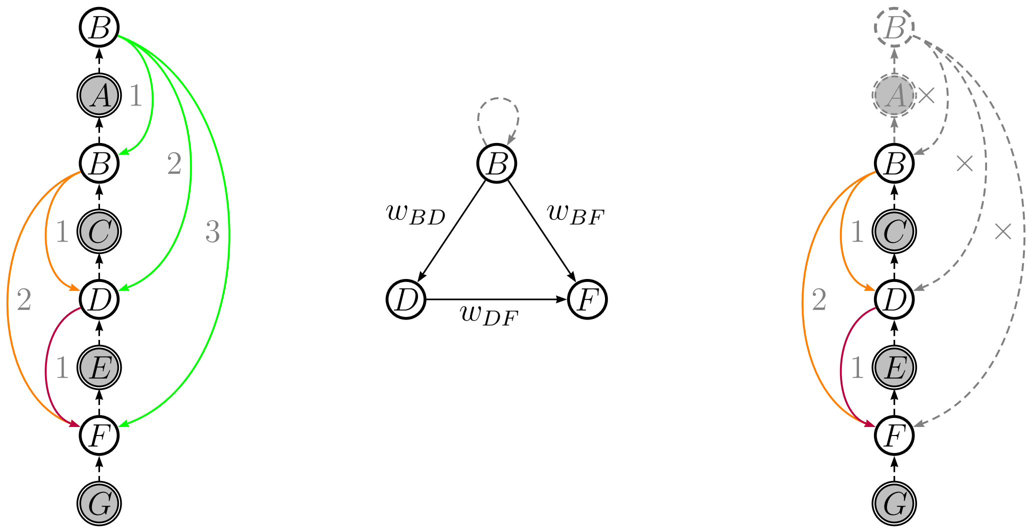
$$w_{DF} = \frac{\sum\limits_{i_k \in \mathfrak{I}_{DF}} \left(\dfrac{1}{i_k}\right)}{|\mathfrak{I}_{D*}|} = \frac{\frac{1}{1}+\frac{1}{1}}{2} = \frac{2}{2} \qquad w_{BH} = \frac{\sum\limits_{i_k \in \mathfrak{I}_{BH}} \left(\dfrac{1}{i_k}\right)}{|\mathfrak{I}_{B*}|} = \frac{\frac{1}{1}}{4} = \frac{1}{4}$$

Notice that though argument F always follows after D, $w_{BF} < w_{BD}$.

The intuition behind this approach draws from the study of electricity. If one assumes that the θ-distance between two arguments in a dialogue acts as a metric in an analogous way to that of resistance in an electric circuit, then using that resistance as the denominator in a fraction where the denominator is 1 resembles the notion of conductivity in a circuit. In our case this conductivity is expressed in the form of $\frac{1}{i_k}$ while the sum of these conductivities serves as an indicator of how easy it is to get to a certain argument from a certain point. Once again, the use of the denominator $|\mathfrak{I}_{A*}|$ serves only for normalisation purposes.

5.2.1 Comparing the two weighing mechanisms

At this point it is worth comparing the two weighing mechanisms proposed in this thesis. To recap, assuming a relationship between two arguments A and B, according to the mechanism described in Definition 52, we simply count the number of times that argument A appears followed by B, and we divide that number with the total number times that A generally appears in dialogues followed by any argument. In essence, Definition 60 is an extension of Definition 52, as it employs the same basic approach. Namely, to count the instances of the, allegedly, related arguments in dialogues and normalise that number in order to serve as a probability value. The only difference is that instead of simply caring



(a) (b) (c)

Figure 5.5: a) A dialogue between two agents (we assume that the grey agent is the modeller), b) the induced $\mathcal{R}G$, c) ignoring A's attack

about whether two opponent arguments, one subsequent to the other, have just appeared in a dialogue, we instead record their appearances in discrete disputes, as well as the distances they have from each other.

We should note that, given the protocol restrictions that characterise either the credulous or the grounded dialogue games, it is not possible for the same pair of related arguments to appear twice in the same dispute, since either the proponent or the opponent will be restricted from repeating an argument they have previously used in the concerned dispute. It is however possible for the same opponent argument to appear twice in the same dispute. Let us consider for example a sequence of arguments moved by two agents in a game that appears in Figure 5.5. Statistically, the appearance of B after itself simply means that it is likely for B to follow after itself in a dialogue (the dashed loop on B in Figure 5.5b). However, semantically this is not really useful since if B is moved into the game then the modeller already becomes aware of the opponent's ability to repeat B contingent on the protocol restrictions and thus accounting for this in its modelling becomes simply redundant.

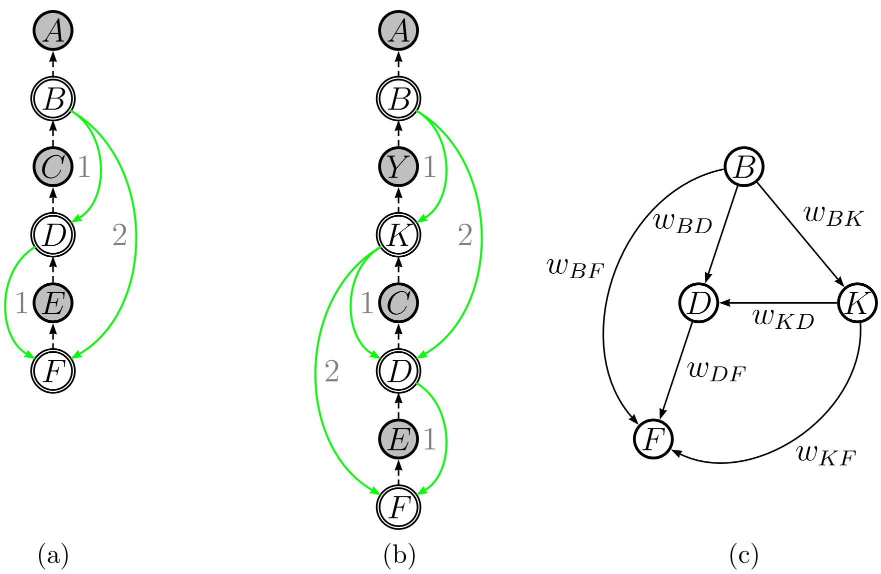


Figure 5.6: a) A dialogue between two agents Ag_1 and Ag_2, b) Another dialogue between Ag_1 and Ag_3, c) The induced $\mathcal{R}G$ for a $\theta_t = 2$

Nevertheless, the repetition of the opponent argument B here can still be recorded without affecting the validity of our statistical approach. Of course, depending on the value of θ_t the argument distances from B will vary, e.g. argument D will appear at θ-distances $\{1, 2\}$ from B (Figure 5.5a), which means that its relationship likelihood with B, particularly in the case of the second weighing mechanism, would actually decrease compared to a case where B would not be repeated (Figure 5.5c). But this will be the case with all other arguments related to B as well. One may, however, choose to simply ignore its repetition, i.e. to ignore A's attack on B, converting such disputes into their semantically equivalent forms, such as the conversion of Figure 5.5a to Figure 5.5c, and apply the weighing mechanisms on those.

In order to better understand the differences between the two mechanisms we go through a trivial example of building a $\mathcal{R}G$ and assigning weight values to its arcs, computing these values by applying both of them.

Example 9 *Let Ag_1 and Ag_2 engage in a dialogue game depicted in Figure 5.6a where we assume that Ag_1 is the modeller (the grey nodes). Assume that Ag_1 engages in another dialogue with Ag_3 depicted in Figure 5.6b. The curved green* *arrows that appear in the two dialogues indicate the θ-distances between the opponent arguments that appear, and these distances are bounded at a $\theta_t = 2$, while the $\mathcal{RG}$ induced from the two dialogues appears in Figure 5.6c.*

Assigning the weight values on the arcs of the induced $\mathcal{RG}$ can be done either by relying on Definition 52 or Definition 60. Table 5.1 displays the computation as well as the produced values for all the weights for both cases.

	$w_{AB} \equiv \frac{M_{AB}}{M_{A*}}$	$w_{AB} = \frac{\sum_{i_k \in \mathfrak{I}_{AB}} \left(\frac{1}{i_k}\right)}{\lvert\mathfrak{I}_{A*}\rvert}$
w_{BF}	$\frac{M_{BF}}{M_{B*}} = \frac{1}{2} = 0.5$	$\frac{\sum_{i_k \in \mathfrak{I}_{BF}} \left(\frac{1}{i_k}\right)}{\lvert\mathfrak{I}_{B*}\rvert} = \frac{\frac{1}{2}}{2} = 0.25$
w_{BD}	$\frac{M_{BD}}{M_{B*}} = \frac{2}{2} = 1$	$\frac{\sum_{i_k \in \mathfrak{I}_{BD}} \left(\frac{1}{i_k}\right)}{\lvert\mathfrak{I}_{B*}\rvert} = \frac{\frac{1}{1}+\frac{1}{2}}{2} = 0.75$
w_{BK}	$\frac{M_{BK}}{M_{B*}} = \frac{1}{2} = 0.5$	$\frac{\sum_{i_k \in \mathfrak{I}_{BK}} \left(\frac{1}{i_k}\right)}{\lvert\mathfrak{I}_{B*}\rvert} = \frac{\frac{1}{1}}{2} = 0.5$
w_{DF}	$\frac{M_{DF}}{M_{D*}} = \frac{2}{2} = 1$	$\frac{\sum_{i_k \in \mathfrak{I}_{DF}} \left(\frac{1}{i_k}\right)}{\lvert\mathfrak{I}_{D*}\rvert} = \frac{\frac{1}{1}+\frac{1}{1}}{2} = 1$
w_{KD}	$\frac{M_{KD}}{M_{K*}} = \frac{1}{1} = 1$	$\frac{\sum_{i_k \in \mathfrak{I}_{KD}} \left(\frac{1}{i_k}\right)}{\lvert\mathfrak{I}_{K*}\rvert} = \frac{\frac{1}{1}}{1} = 1$
w_{KF}	$\frac{M_{KF}}{M_{K*}} = \frac{1}{1} = 1$	$\frac{\sum_{i_k \in \mathfrak{I}_{KF}} \left(\frac{1}{i_k}\right)}{\lvert\mathfrak{I}_{K*}\rvert} = \frac{\frac{1}{2}}{1} = 0.5$

Table 5.1: The weight values of the arcs of the $\mathcal{RG}$ of Example 9 produced by the two weighing mechanisms

One can immediately notice the differences between the two weighing mechanisms, concluding that in the second case the weight values are generally lower, which is specifically evident in the first and the second row as well as in the last row of Table 5.1. This appeals to our intuition as lowering the likelihood of two arguments to be related is a reasonable thing to do if those arguments are not directly supported. Of course, as this is a trivial example concerned with just two dialogues, the calculated weights are far from accurately representing the actual relationship likelihood between their concerned linked arguments due to the cold start problem. However, the example is enough to illustrate the differences between the two approaches.

Let us further elaborate on this issue by focussing on a certain case of the presented example. In the case of weight w_{BD} and assuming application of the first mechanism the number of times that the pair (B, D) appeared is $M_{BF} = 2$ while the total times that argument B was used is $M_{B*} = 2$, resulting in a weight value equal to $w_{BD} = 1$. In contrast, if we apply the second mechanism then since the pair appears twice at distances $i_1 = 1$ and $i_2 = 2$ we have a sum of their conductivities equal to $\frac{1}{1} + \frac{1}{2} = 1.5$, which is then divided by the instances of B which are equal to $|\mathcal{I}_{B*}| = 2$ to produce a weight value $w_{BD} = 0.75$.

It is thus evident that in the second case the weight values are lower than in the first case. More interestingly though, if we attempt to increase the θ_t bound to 3 thus allowing arguments B and F to be assumed related both in the first as well as in the second dialogue (while with a $\theta_t = 2$ they were just related in the first dialogue), assuming application of the second weighing mechanism will result in an increase of its weight value, though a minor one. Specifically, for a $\theta_t = 2$:

$$w_{BF} = \frac{\frac{1}{2}}{2} = 0.25$$

for a $\theta_t = 3$, it will become equal to:

$$w_{BF} = \frac{\frac{1}{2} + \frac{1}{3}}{2} = 0.44$$

In general, this approach enhances the possibility of two indirectly supported arguments to be related as the θ_t bound increases. We finally note that, the choice of weighing mechanism depends on the user's preferences and the domain of the application.

5.3 Common Attack Targets–$\mathcal{RG}$s

So far we have stressed that for linking arguments in a $\mathcal{R}G$ one may rely on inherent properties between the OAs used in a modeller's history of dialogues. However, we have only used the notion of support for doing so, i.e. the arguments' ability to reinstate one another. Other notions can be used to serve for expressing different kinds of interrelatedness of information to link arguments in an $\mathcal{R}G$. One such notion is that of arguments with *common attack targets*. In other words, to assume that it is possible for someone capable of attacking an argument A with an argument B, to be aware of additional arguments (e.g. C, D, E) that can do so as well (attack A).

The modelling approach in this case is very similar to the one that relies on the notion of support, where we essentially assume that opponent arguments in one's general history of dialogues that appear to reinstate other arguments that are already part of an OM, could, with some likelihood, also be part of that opponent's knowledge. In an analogous way, appearance of certain opponent arguments usually used against a modeller's certain moves in that modeller's general history of dialogues, could be used to anticipate other arguments that might be known to that opponent, given the fact that they attack the same target.

Let us present an intuitive example. Assume that someone is trying to persuade a member of the Scottish National Party (SNP) that Scotland should stay in the union (claim q) because it is in their economic interest (premise p). Let us refer to this argument as A. To this argument the member of the SNP offers a counter argument B stating that Scotland should exit the union (claim $\neg q$), since the reason for originally joining the union no longer applies (premise w). Assume however that another three counter-arguments could have been presented: C: also stating that Scotland should exit the union (claim $\neg q$), because Scottish people are generally different than English; D: stating that it is not in Scotland's economic interest to remain in the union (claim $\neg p$), because the monetary policy of the United Kingdom is biased towards England (premise t), and E: also stating that it is not in Scotland's economic interest to remain in the union (claim $\neg p$), because natural resources generated wealth would be better distributed among the Scottish rather than among all of the United Kingdom citizens (premise r).

One could assume that arguments B, C, D and E are allegedly associated with some likelihood. In order to allow for their interrelatedness to become clearer let us investigate their logical structure. We have the first argument:

$$A : p; p \Rightarrow q$$

attacked by another four arguments:

$$B : w; w \Rightarrow \neg q, \quad C : s; s \Rightarrow \neg q, \quad D : t; t \Rightarrow \neg p, \ and \ \ E : r; r \Rightarrow \neg p$$

One can observe that arguments B and C produce the same claim ($\neg q$), which also holds for arguments D and E ($\neg p$). Thus the relationship between the elements of the two pairs is evident. As for the relationship between the pairs themselves, it lies in the fact that they both share contrary or contradictory relationships with A—the constituents of which are p and q. The extent to which one should assume that the elements of this set of arguments ($\{B, C, D, E\}$) are related, derives from how often they appear together to attack argument A in the same dialogue.

Specifically, we will be linking arguments in a $\mathcal{RG}$ if they appear in distinct disputes, though in the same dialogue, following directly after a certain proponent argument, i.e. attacking the same target in a dialogue. Example 10 illustrates this modelling approach.

Example 10 *Let Ag_1 engage in a dialogue with Ag_2 as shown in Figure 5.7a, followed by another dialogue with another agent (Ag_3) as it appears in Figure 5.7b, and a third one with another agent (Ag_4) which appears in Figure 5.7c.*

If we assume the construction of a $\mathcal{RG}$ based on the common attack targets notion, then, based on the first dialogue, arguments B, C, and D will be all linked with each other as they attack argument A. As no other proponent argument is attacked by more than one argument, no other relationships can be formed based on the first dialogue. Another two links can be established between arguments M and H based on the second dialogue as they both attack argument E. Finally, the third dialogue contributes to the establishment of another pair of links between

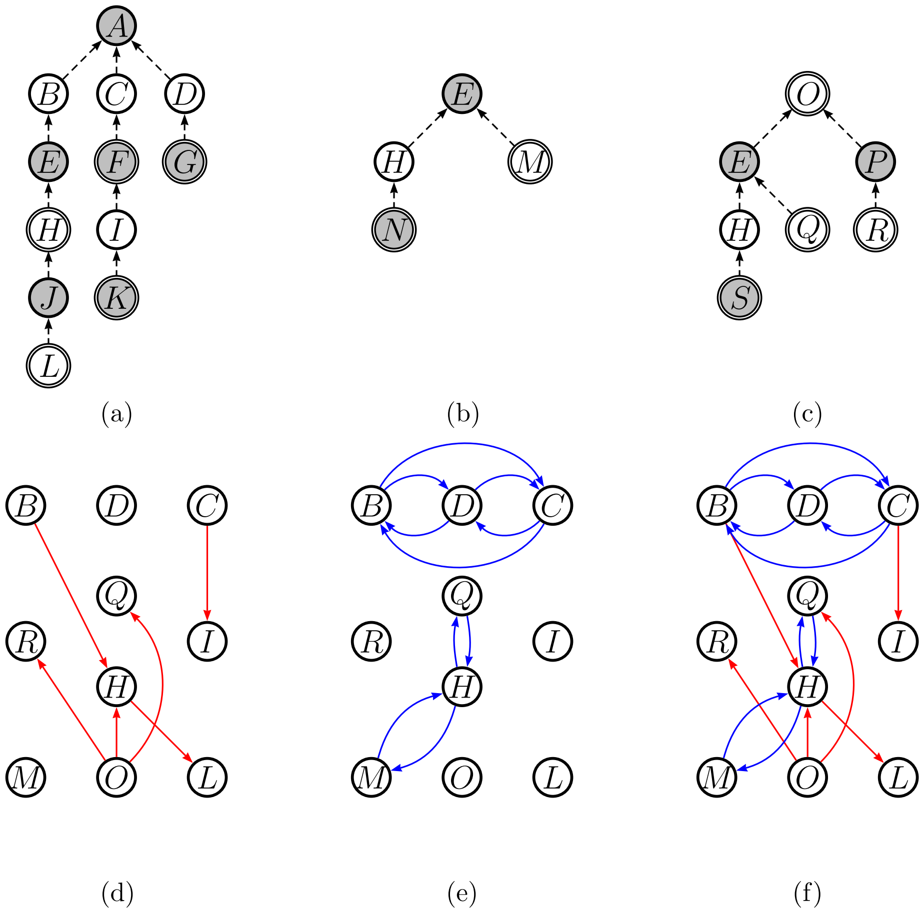


Figure 5.7: a) A dialogue between Ag_1 and Ag_2, b) A dialogue between Ag_1 and Ag_3, c) A dialogue between Ag_1 and Ag_4, d) A $\mathcal{RG}$ induced from the three dialogues based on the notion of support, e) An $\mathcal{RG}$ induced from the three dialogues based on the notion of common attack targets, f) The combined result of the two $\mathcal{RG}$s

arguments H and Q as they both attack E, while arguments R, O, I and L will be simply added in the $\mathcal{R}G$ with no links at all. The resulting $\mathcal{R}G$ appears in Figure 5.7e.

Notice that in contrast with relying on the support notion, the links in the case of arguments with common attack targets are reciprocal. Take for example two arguments A and B with a common attack argument C, it makes little sense to assume that awareness of A implies a likely awareness of B but not vice versa. Also notice that, based on the same dialogue restriction, even though one would expect, arguments Q and M to also be linked as they both attack E in Example 10, they do not. This is because our objective is to monitor and model a relationship between arguments *in the same dialogue*, as we are specifically interested with anticipating whether an argument will follow after another in a dialogue or not. In this sense, appearance of an attacker along with other attackers in the same dialogue is essential.

Definition 61 provides the conditions that must be satisfied for the establishment of a link between two OAs in a $\mathcal{R}G$.

Definition 61 (Connectivity Condition II) *Let d' be a sub-dispute of a dispute $d \in \mathcal{T}$ such that $d' =< \mathcal{DM}_0, \ldots, \mathcal{DM}_k >$, and let A and B be two arguments respectively serving as the content of two opponent dialogue moves $\mathcal{DM}_i$ and $\mathcal{DM}_j$ in a dialogue tree $\mathcal{T}$. Then if both $\mathcal{DM}_i$ and $\mathcal{DM}_j$ extend d' in $\mathcal{T}$ such that:*

- *$\exists d'_1 =< \mathcal{DM}_0, \ldots, \mathcal{DM}_k, \mathcal{DM}_i >$, where d'_1 is a sub-dispute of a $d_1 \in \mathcal{T}$, and;*
- *$\exists d'_2 =< \mathcal{DM}_0, \ldots, \mathcal{DM}_k, \mathcal{DM}_j >$, where d'_2 is a sub-dispute of a $d_2 \in \mathcal{T}$*

then $\exists r_{AB} \in \mathcal{R}$ and $\exists r_{BA} \in \mathcal{R}$.

For computing the weight values on the induced arcs we simply count the number of times that a certain pair of linked arguments appeared in the same dialogue. This number can serve as the numerator of a fraction which will eventually represent the desired weight value. As with the weights assigned on the support relationships, these values also need to be normalised so as to express probability—likelihood—values. For this purpose we can use two possible numbers as denominators. These are, assuming a pair of linked arguments A and B, the number of times that A appeared in dialogues, or; the number of times that B appeared in dialogues. Both numbers will reasonably be greater than the number of the respective appearances of their pairs resulting in bounding a weight value in the closed range of $[0, 1]$. However, instead of resorting to the use of one of the two for assigning equal weights on the two edges between the two arguments, we can use the number of times that A appears in dialogues as the denominator for the r_{AB} edge, and the number of times that B appears as the denominator for the r_{BA} edge. In other words, if we assume that the weight value of r_{AB} is a response to the question: *How many times has A appeared in dialogues with B, both attacking the same argument?*, then that weight should be equal to the number of times that argument A appeared in dialogues with B, attacking the same argument, divided by the total appearances of A, and vice versa for r_{BA}.

Definition 62 formally expresses this approach.

Definition 62 (Weight Assignation II) *Let $\mathcal{R}G = \{\mathcal{A}^{\mathcal{H}}, R\}$ be an agent's relationship graph, while arguments A, B and C are elements of $\mathcal{A}^{\mathcal{H}}$ then:*

$$\texttt{Instances}(\mathcal{H}, A, B, C) = N_{AB}$$

is a function that returns a number N_{AB} representing the number of times that arguments A and B attacked argument C in separate disputes in the same dialogue in $\mathcal{H}$, such that:

$$\exists d_1' \& =< \mathcal{DM}_0, \ldots, \mathcal{DM}_k, \mathcal{DM}_i >, \quad \textbf{\textit{and}};$$
$$\exists d_2' \& =< \mathcal{DM}_0, \ldots, \mathcal{DM}_k, \mathcal{DM}_j >$$

where d_1' is a sub-dispute of d_1, and d_2' is a sub-dispute of d_2, for $d_1, d_2 \in \mathcal{T}$ and where $\mathcal{DM}_i^{con} = A$, $\mathcal{DM}_j^{con} = B$ and $\mathcal{DM}_k^{con} = C$. Then:

$$w_{AB} \equiv N_{AB}/N_{A*}$$
$$w_{BA} \equiv N_{BA}/N_{B*}$$

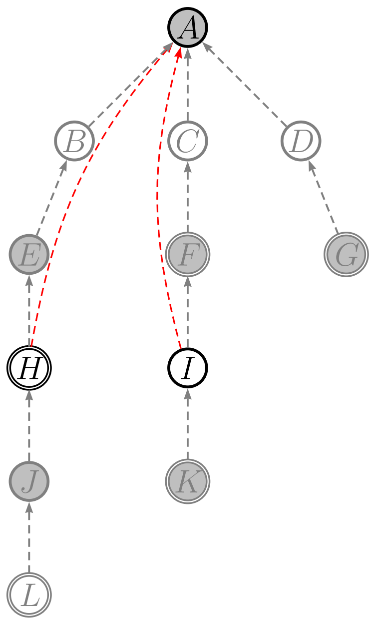


Figure 5.8: Arguments H and I indirectly attacking the same argument, A

Obviously, in this case as well, the cold start problem still remains. Furthermore, the weighing mechanism proposed in Definition 62 as well as the linking condition of Definition 61 could both be extended in a similar way to Definitions 51 and 52, to model an indirect common attack target relationship. Take, for example, arguments H and I in Figure 5.8. One could assume that in contrast with arguments B, C and D which are direct attackers of A and are thus assumed related, H and I could also be related as they are indirect attackers of A. Furthermore, in addition to the relationships between all the attackers—direct or indirect—they could all be associated in various ways, with their likelihoods quantified so as to reflect their attacking distance from their target and their general distance from each other. However, this is something we leave for future research.

Finally, it is worth mentioning that instead of opting between the two modelling approaches—the one that relies on the notion of support and the one that relies on the notion of common attack targets—one may choose to apply both. This is described in Figure 5.7f where the edges corresponding to the respective approaches are differentiated with red and blue edges. Notice that in contrast with Figures 5.7d, and 5.7e, Figure 5.7f is more dense, linking all arguments into a single coherent graph. We also note that the significance of a certain approach, according to the user's preferences, may be reflected by appropriate normalisations of the two possible weights. For example, the modeller may multiply all weights produced by either approach with a certain percentage in order to enhance the impact of the approach she favours.

5.4 Conclusions

In this chapter we extended the simple modelling approach presented in Chapter 4. We discussed the added value of modelling $\mathcal{R}G$s with a θ_t bound greater than 1, and we extended the simple weighing mechanism used for quantifying the support relationships in a $\mathcal{R}G$ so as to better reflect that relationship, accounting for the θ-distance between the related arguments. In other words, we proposed a way for modelling indirect support relationships between arguments.

We also showed how building $\mathcal{R}G$s may rely on other inherent properties that relate arguments apart from that of support. We proposed a linking condition that relies on the notion of arguments with common attack targets and provided a corresponding weighing mechanism for the quantification of those relationships.

Our main intention was to illustrate the multifaceted way based on which one may construct an $\mathcal{R}G$. We showed that by simply relying on inherent linking properties between arguments it is easy to associate them and quantify their relationship by monitoring patterns of their appearances in an agent's general dialogue history.

There are several other issues drawn from our work which are worth researching. One such issue is the construction of $\mathcal{R}G$s with the logical constituents of associated arguments, rather than with just arguments, which would allow for a better understanding of the interrelatedness of information. Also, investigating the differences between expanding/augmenting one's knowledge on the fly or offline, is also an interesting problem to investigate. In other words, to research the differences between the resulting gained knowledge and the advantages of using the one or the other. We intend to address such problems in future work.

Chapter 6

On the Use of Confidence Values in Strategising

Dealing with the best response problem in competitive game contexts where the objective is to achieve the maximum utility from a certain choice, usually orients around von Neumann's [1928] minimax theory. As application of the minimax algorithm is conditional on the modeller's knowledge of the possible knowledge of its opponents, it is imperative that we develop mechanisms able to anticipate what could be known to those opponents which we presented in Chapter 4. This allows for the construction of a *game tree* that simulates the ways based on which a dialogue may evolve. The basic steps that follow after construction of such a tree are two:

1. Application of a *UEF* for estimating the modeller's utility of the game possibly ending at each of the tree's leafs, and;

2. Application of the *minimax algorithm*.

In this chapter we develop and present a general UEF used for the evaluation of the leafs of a game tree. We illustrate how this function accounts for confidence values assigned to the possible opponent choices (arguments) that represent the likelihood of an argument being actually known to an opponent. We then elaborate on how the minimax algorithm may be applied on the evaluated game tree in order for a decision to be made at a strategic point.

6.1 Introduction

In Chapter 4 we explored issues related to modelling an opponent's knowledge base. According to the literature, most approaches related to opponent modelling techniques are mostly relying on mechanisms the definition of which was usually left implicit. Our contribution towards bridging this gap is the development of three mechanisms which respectively relied on three different information collection methods (ICMs) which could be used for opponent modelling purposes. In addition, for each of these methods we developed a confidence assignation function for deciding on the likelihood according to which a modeller should believe that its opponent's are indeed aware of certain information. Our interest in this chapter is to illustrate how exactly these confidence values can be used for strategic reasoning purposes, through the combined use of a UEF and the application of the minimax algorithm.

The success of the minimax algorithm relies on two main factors: the accuracy of the UEF, and; the assumption that the possible outcomes for the two participants of the game are inversely dependent, i.e. that one's gain is the other's loss, which, when combined with the assumption that the opponent is a reasonable agent, leads to the conclusion that when presented with a choice the opponent will attempt to also maximise its utility and, ergo, to minimise the proponent's.

However, the assumptions on which the minimax algorithm relies, are often debated. For one, assuming that one's opponent will apply the exact same decision making processes, suggests that it is easy to anticipate the actions of that opponent, and thus to counter-strategise. Arguably, when strategising one has to also account for this possibility since otherwise the effectiveness of an employed strategy can be considerably diminished. However, in most cases, the run-time of computing the presumingly optimal response at a given state in a game, through additionally accounting for counter-strategising, drastically increases, deeming most theoretical approaches practically not tractable.

Approaches that attempt to account for counter-strategising are presented in the work of Carmel and Markovitch [1996] and Oren and Norman [2010] which seem to be building on ideas originally proposed by Korf [1989]. In both cases nested opponent models are proposed in the form of a recursive definition, drawing from the work of Gmytrasiewicz et al. [1991]. We should clarify that these models refer to the possible strategies which could be employed by one's opponent, and not to its knowledge. Variants of the minimax algorithm are proposed in both cases as well (Carmel and Markovitch [1996]; Oren and Norman [2010]), while pruning techniques are also provided that reduce run-time by decreasing the search space, achieving a feasible approximative computation of the optimal response.

An interesting approach is proposed by Rienstra et al. [2013] where a recursive structure is used for modelling the beliefs an agent has on another agent's beliefs through the use of *virtual arguments*, i.e. arguments which could possibly counter a modeller's strategy if known by its opponent. Their approach relies on synthesising quantitative and qualitative uncertainty information concerned with both the actual OM and the set of the believed arguments respectively for strategising. They were able to show, through experimenting, that the increased complexity of the opponent modelling structure results in improving the strategy outcomes for the modeller.

At this point it is worth mentioning the work of de Weerd et al. [2013], titled *"How much does it help to know what she knows you know? An agent based simulation study"*. The researchers of this work performed a series of experiments based on four competitive games, where the participants were rational agents with different orders (meta-levels) of cognitive ability, to which they refer as *"theory of mind"*. To their surprise, though both first-order and second-order theory of mind agents were able to outperform adversaries with limited abilities, they found that higher order of theory of mind shows diminishing returns. They generally conclude that resorting to the use of the ability of the theory of mind provides an advantage when employed in complex competitive game settings. In the context of simple games participants seem to be able to benefit from simply monitoring their opponents actions in previous interactions (i.e. through recording their actions' history).

Though dealing with issues like counter-strategising and nested opponent modelling is an inherent aspect of strategising in dialogues, they are not in the scope of our work. However, for the sake of theoretical completeness, we refer the reader to Section A.2 of Appendix A, where we show how our model can be extended to incorporate the notion of nested OMs (S'), where, in contrast to the work of Carmel and Markovitch [1996], and Oren and Norman [2010] we model opponent knowledge instead of opponent strategies.

Another issue concerns the impact of the UEF on the result of the minimax algorithm. The minimax algorithm is classically applied on zero-sum games, and persuasion dialogue games are such games, since they can either result in a win or a loss. However, victory and defeat are not the only factors accounted by a UEF. Secondary objectives are also accounted such as the possible information which could be gained through a certain dispute, or the extent to which the dialogue was stalled etc. Thus, for assuming that one's opponent will apply the minimax algorithm for strategising, the modeller builds on the assumption that she can anticipate exactly how the opponent will evaluate the possible outcomes of the game. In other words, the modeller assumes to know the opponent's objectives for deducing the opponent's possible strategy.

However, a participant's strategy cannot be anticipated prior to a game, since the participants goals are not always evident. Though, one would expect agents to comply to certain dialogical commitments imposed by the nature of dialogues, as McBurney and Parsons [2002] explain in their work, agents may deviate from those commitments, sometimes even halfway through a game. However, it may be to some extent possible to deduce one's objectives during a game, by monitoring its actions while playing. This is also suggested in the work of Oren and Norman [2010]. As they explain, provided the knowledge about an opponent's goals it is also possible to indirectly model its strategy. However, in games of imperfect information, deducing one's goals during a game can be a tricky problem.

To the best of our knowledge, these are interesting issues which still remain unresolved in the fields of Game Theory and Artificial Intelligence. Thus, for the time being, we can only focus on factors which we can directly affect. Such factors concern the accuracy of the UEF applied on the leaf-nodes of a game tree, which in turn relies on the validity of the information encapsulated in an OM. If we can develop techniques, relying on the interrelatedness of information, for collecting credible information about the possible knowledge of one's opponents, then we can increase the effectiveness of any strategy that will be relying on them. This was our intention in Chapters 4 and 5. In this chapter we focus on exploring how we can use the collected opponent information for strategising.

The rest of this chapter is structured as follows: in Section 6.2 we provide a definition for a utility evaluation function which uses the confidence values assigned to arguments instantiated from an OM, and show how it is applied on a *game tree* derived from a single or from multiple dialogue trees; in Section 6.3 we elaborate on the consequent application of the minimax algorithm on that game tree, which leads to making a decision in a dialogue when a modeller is found at a strategic point; we finally summarise our work in Section 5.4.

6.2 Utility Evaluation

In Chapter 4 we offered a definition for associating all the logical elements in an OM ($S_{(i,j)}$) with a confidence value c (Definition 46). To recap, assuming an OM $S_{(i,j)}$ then for every set Y where $Y \in \{\mathcal{K}_{(i,j)}, \leq'_{(i,j)}, \mathcal{R}_{(i,j)}, \leq_{(i,j)}, \mathcal{G}_{(i,j)}\}$ we assume tuples X of the form $< x, c >$ where $x \in Y$ and c is a numerical value in the range of 0 and 1 which is associated with x and which depends on the ICM used for the inclusion of x in $S_{(i,j)}$. Then, for determining the confidence value c_A of an opponent argument (OA) we choose to rely on Definition 58 according to which an argument A instantiated from a set of logical constituents $\Lambda = \{x_1, x_2, \ldots, x_i, \ldots, x_n\}$ is assigned the lowest confidence value associated to those logical constituents, i.e.:

$$c_A = min\{c_1, c_2, \ldots, c_n\}$$

This section elaborates on the provision of a UEF which utilises these confidence values for assigning a utility value to each of the anticipated outcomes of a dialogue. Having provided a method for determining the likelihoods of a set of OAs to actually be known to a certain opponent, we can then utilise them for computing the utility of a possible outcome of a dialogue game, i.e. the utility of a leaf of a game tree.

6.2.1 Game Trees

Generally, a utility evaluation function is applied to each leaf-node of a *game tree*, assigning a numerical utility value to them, which represents how good each node is as an outcome for the modeller. However, application of a UEF is not possible, or precisely meaningless, on dialogue trees as they differ from game trees in two important aspects.

Firstly, a dialogue does not necessarily end at a leaf-node of a dialogue tree; due to the possibility of backtracking. In other words, since a leaf-node in a dialogue tree does not represent a terminal state in a dialogue game it cannot be evaluated. Let us consider the example in Figure 6.1. We note that in contrast to the notation used for the representation of dialogue trees, arrows in a game tree will have the opposite direction, i.e. if B attacks A in a dialogue tree ($B \rightarrow A$) that means that it follows after A in a game tree ($A \rightarrow B$). In addition, double edged arrows will be used in game trees as opposed to dashed arrows in dialogue trees, while all nodes in a game tree will be represented with a single edge line. Fill colours grey and white are maintained for differentiating between the arguments used by the modeller and its opponent respectively.

Figure 6.1a illustrates a dialogue tree where two leaf-nodes appear: D and E. However, since E was submitted on backtracking, only E is the terminal node. Let Figures 6.1b & 6.1c represent the possible ways based on which the dialogue may evolve. Figure 6.1c depicts the path followed in the case where when found at the strategic point of choosing between arguments C and E the modeller opts for E (possible path II). Alternatively, Figure 6.1b depicts the case where C is chosen (possible path I). Path I, in contrast to path II, includes a backtrack move which follows after D, since according to the dialogue tree (Figure 6.1a), there is no other argument able to attack D, and E is offered as the only alternative. The combination of the two possible paths (I & II) produces the game tree $\mathcal{GT}$ that appears in Figure 6.1d. Obviously, in the case where backtracking is not allowed, the dialogue tree and the game tree will be the same with the latter simply having its arrows turned to the opposite direction.

The second difference between the two trees is that though they are both simulations of the possible ways based on which a dialogue may evolve, a dialogue

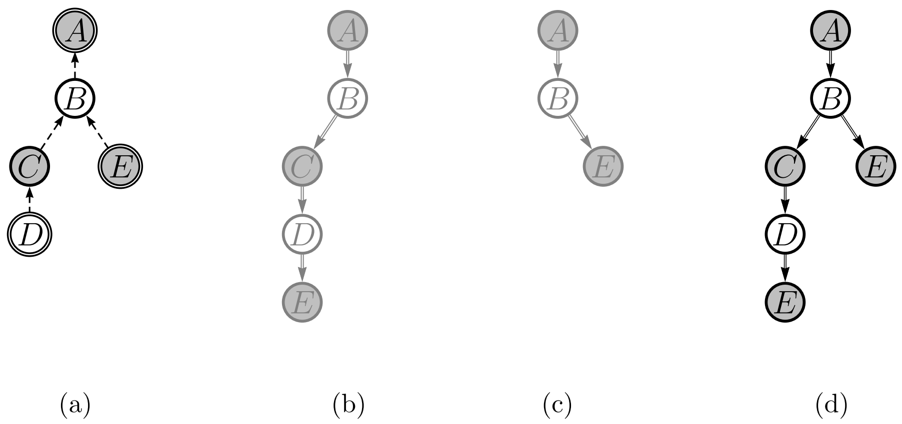


Figure 6.1: Conversion of a dialogue tree to a game tree: a) A dialogue tree $\mathcal{T}$, b) Possible path I, c) Possible path II, d) The resulting game tree $\mathcal{GT}$

tree can only encapsulate a subset of those possibilities, while a game tree encapsulates all of them. Let us reconsider the example of Figure 6.1, this time from the perspective of the modeller. Let us assume that the modeller (the proponent) is aware of arguments:

$$\mathcal{A}_{(Pr,Pr)} = \{A, C, E, F\}$$

while he assumes that his opponent is aware of arguments:

$$\mathcal{A}_{(Pr,Op)} = \{B, D\}$$

while let `attacks` represent the binary attack relationships between them, then:

$$(B, A), (C, B), (D, C), (E, B), (F, B) \in \texttt{attacks}$$

In this case, if the modeller attempts to simulate the possible ways based on which the dialogue may evolve he would face a dilemma: After backtracking from Op's move D, then out of the two possible options E or F, which should she use to counter B (Figure 6.2a)? Obviously, using both is not possible as it would violate game protocol restrictions according to which every agent can move a single argument at a time. It is therefore evident that for simulating all possible ways based on which the game could evolve, the modeller has to construct two dialogue trees, $\mathcal{T}_1$ and $\mathcal{T}_2$ as they appear in Figures 6.2b and 6.2c.

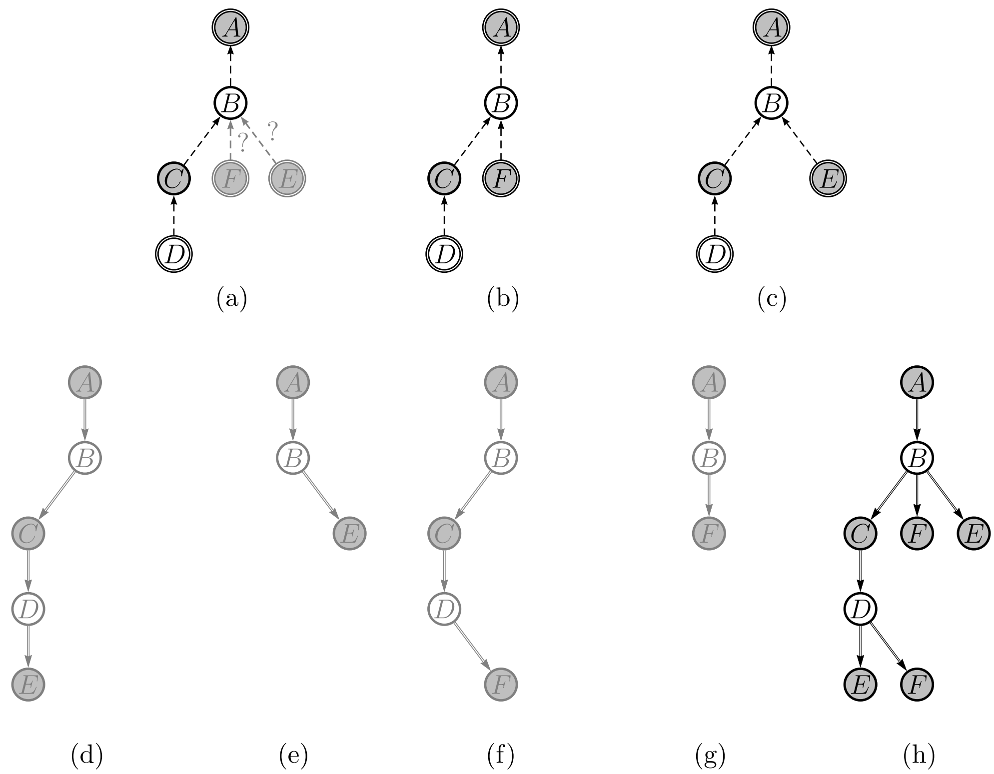


Figure 6.2: a) An incomplete dialogue tree where only one between two options can be used to complete it, b) Possible dialogue tree $\mathcal{T}_1$, c) Possible dialogue tree $\mathcal{T}_2$, d) Possible path I, e) Possible path II, f) Possible path III, g) Possible path IV, and h) the complete game tree.

As already noted, these two trees express two discrete subsets of the possible ways based on which the dialogue may evolve, the complete set of which is the sum of Figures 6.2d, 6.2e, 6.2f and 6.2g, which is expressed in the the form of game tree in Figure 6.2h.

In sum, for the complete construction of a game tree a modeller has to go through three steps:

1. instantiate all the possible dialogue trees, i.e. the forest of dialogue trees, that can be constructed based on his own knowledge and his assumptions about his opponent's knowledge;

2. produce their corresponding game trees, and;

3. merge these game trees in order to form a complete game tree.

We note that initial construction of the possible dialogue trees is necessary, as only dialogue trees adhere to the protocol restrictions imposed on the possible moves introduced by the participants at each point of a dialogue.

In this respect, we provide a formal description of a dialogue tree's corresponding game tree ($\mathcal{GT}$) in Definition 65. To do so we first formally define a dialogue tree's reply structure, as well as the basic properties of a game tree in Definitions 63 and 64 respectively. At this point we advise that the reader recalls Definition 31 in which $\mathcal{E}$ represents the set of edges between the nodes of a dialogue tree, and $\mathcal{DM}_r$ is the dialogue tree's root dialogue move.

Definition 63 (Reply Structure) *Given a dialogue tree $\mathcal{T}$ we assume a set $\mathscr{R}$ that represents its reply structure:*

$$\mathscr{R} = \{(X, Y) | \exists (\mathcal{DM}_i, \mathcal{DM}_j) \in \mathcal{E} : \mathcal{DM}_j^{con} = X, \mathcal{DM}_i^{con} = Y\}$$

where X and Y are arguments.

In simple words, Definition 63 states that for every pair of arguments $(X, Y) \in \mathscr{R}$ there exists a corresponding pair of dialogue moves $(\mathcal{DM}_i, \mathcal{DM}_j) \in \mathcal{E}$ whose respective contents are arguments Y and X, where X attacks Y. We note that a reply structure differs from the set of attack relationships between the arguments employed in a dialogue game, as there may be reciprocal attacks between arguments which cannot appear in the dialogue game due to protocol restrictions.

Provided a dialogue tree's reply structure we define a game tree as follows:

Definition 64 (Game Tree) *A game tree $\mathcal{GT}$ is an acyclic graph expressed as a tuple of the form $\mathcal{GT} =< \Sigma N, \mathcal{P} >$ where $\Sigma N = \{N_0, N_1, \ldots, N_m\}$ is a set of nodes, expressed as tuples of the form:*

$$N_i =< A, c_A, u_i >$$

where for $i = 0, 1, \ldots, m$:

- *A is an argument;*

- *c_A is that argument's confidence value;*
- *u_i is the utility of that node, and;*

and $\mathcal{P} \subseteq \Sigma N \times \Sigma N$. In addition let:

- *N_0 be the **root node** of $\mathcal{GT}$, while let every other node receive its index value in accordance to a depth-first search;*
- *$L_N = \{N_{l1}, N_{l2}, \dots, N_{lr}\}$ be the set of **leaf-nodes** of $\mathcal{GT}$ every member of which is a terminal node, i.e. it has no children;*
- *$P_{0\to lk}$ be a **path** representing a sequence of nodes $< N_0, \dots, N_{lk} >$, where $k = 1, 2, \dots, r$;*
- *$\Pi = \{P_{0\to l1}, P_{0\to l2}, \dots, P_{0\to lr}\}$ be the set of all possible paths in a $\mathcal{GT}$, and;*
- *$< N_i, N_{i+1} >$ be a subsequent pair of nodes in a $P_{0\to lk}$, where $0 \leq i \leq lk-1$, we then say that N_i and N_{i+1} share a **father** respectively **child** relationship, while every node can have at most one father-node.*

We note that the confidence value for every argument moved by the modeller (that is the proponent in the case of the example depicted in Figure 6.1) is by default assumed to be equal to 1, while for convenience when referring to a single game tree we will be representing its discrete paths using only the index of their leaf-nodes (*e.g.* P_{li} as opposed to $P_{0\to li}$).

Provided Definitions 63 and 64 we define a dialogue tree's corresponding game tree as follows:

Definition 65 (Game Tree Correspondence Criteria) *Given a dialogue tree $\mathcal{T}$, its reply structure $\mathscr{R}$, a game tree $\mathcal{GT}$, and a function **arg()** applied on a node of a $\mathcal{GT}$ to return the argument encapsulated in that node, we then say that $\mathcal{GT}$ is the corresponding game tree of $\mathcal{T}$, iff:*

- *$N_0 =< A, c_A, u_0 >$, where $A = \mathcal{DM}_r^{con}$, and;*
- *$\forall N_j \in P_{0\to lk}$, for $1 \leq j \leq lk$, and $l1 \leq lk \leq lr$, $\exists N_i \in P_{0\to lk}$, where $i < j$ and $0 \leq i \leq j-1$, such that:*

$$\exists(\boldsymbol{arg}(N_j), \boldsymbol{arg}(N_i)) \in \mathscr{R}$$

The first criterion simply states that the root node of a $\mathcal{GT}$ must encapsulate the argument found in the root move of a $\mathcal{T}$. The second criterion states that for every node N_j in a path of a $\mathcal{GT}$, excluding the root node, there exists a preceding node (N_i) in the same path, such that the pair of arguments $(\texttt{arg}(N_j), \texttt{arg}(N_i))$ encapsulated by them exists in $\mathcal{T}$'s reply structure $\mathscr{R}$.

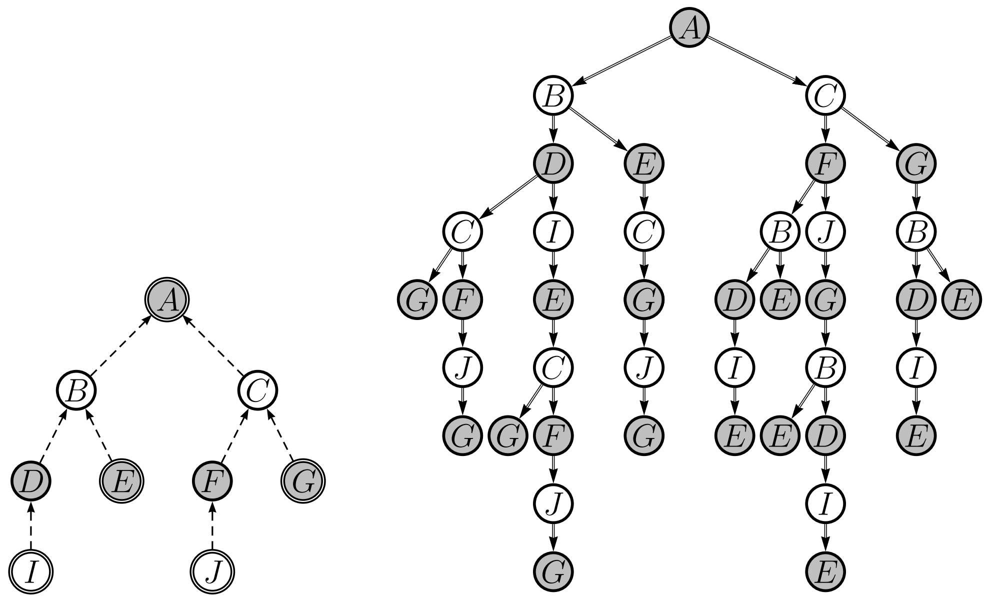
One can easily deduce the relationships between a dialogue tree ($\mathcal{T}$) and a game tree ($\mathcal{GT}$). Namely, a $\mathcal{T}$ is composed of dialogue moves, whereas a game tree of nodes; both have leafs respectively in the form of moves and nodes; while in contrast to a $\mathcal{GT}$ a path between the root move to a leaf move in a $\mathcal{T}$ is referred to as a dispute. Their only essential difference is representational and concerns the fact that directed edges which express the father/child relationships in the two trees are reversed, i.e. in a $\mathcal{T}$ the direction is from the child towards the father while in a $\mathcal{GT}$ is from the father towards the child in an effort to enforce a representational structure able to better reflect the evolution of a dialogue, i.e. which argument/move will follow after another, as opposed to which argument serves as a reply to another.

Consider for example the case of the dialogue tree $\mathcal{T}$ depicted in Figure 6.3a, whose corresponding game tree $\mathcal{GT}$ appears in Figure 6.3b. The reply structure for $\mathcal{T}$ is as follows:

$$\mathscr{R} = \{(B, A), (C, A), (D, B), (E, B), (F, C), (G, C), (I, D), (J, F)\}$$

Let us then take any path of $\mathcal{GT}$, which for convenience will be represented as a sequence of arguments as opposed to the nodes in which they are encapsulated. For example, in the case of path $P =< A, B, E, C, F, J, G >$:

- for G, there is an ancestor node C given that $\exists(G, C) \in \mathscr{R}$;
- for J, there is an ancestor node F given that $\exists(J, F) \in \mathscr{R}$;
- for F, there is an ancestor node C given that $\exists(F, C) \in \mathscr{R}$;
- for C, there is an ancestor node A given that $\exists(C, A) \in \mathscr{R}$;
- for E, there is an ancestor node B given that $\exists(E, B) \in \mathscr{R}$, and;



(a) (b)

Figure 6.3: a) A dialogue tree $\mathcal{T}$, b) its corresponding game tree $\mathcal{GT}$.

- for B, there is an ancestor node A given that $\exists(B, A) \in \mathscr{R}$.

In relation to the technical problem of the actual conversion, we note that it is a complex problem which can otherwise be expressed as finding all possible common-source paths in a graph with multiple destinations. Specification of an algorithm for converting a dialogue tree to a game tree remains a topic for future work.

6.2.1.1 False Assumptions & Logically Imperfect Agents

So far we have accounted for backtracking and its impact in creating additional terminal states in a dialogue's states space, as well as for how one should simulate a forest of dialogue trees in order to account for all the possible ways based on which a dialogue may evolve, as a single dialogue can only represent a subset of those possibilities. However, for instantiating the complete state-space of all possible dialogues between a modeller and an opponent, one has to additionally account for the possibility that a dialogue game may end at any point. In other words, to account for the fact that the participants may not play logically perfect which means that either the proponent or the opponent may opt to not introduce a move into the game, for whatever reason, even though this may result in them losing. Forfeiting a game could be attributed to the respective utility values of a game tree's nodes for either of the participants. It could be the case that the utility of losing the game at a certain point is greater than winning it at another. For example, such a case would arise in a scenario where, based on a simulated game tree, a participant would anticipate certain defeat in later stages of the game in which case revealing more information to the opponent could be detrimental.

In addition, from the modeller's perspective, given that dialogue trees serve only as simulations of the possible ways based on which a dialogue may evolve, then since knowledge assumed to be known by the modeller's opponents *is not certain*, it is possible for the game to end at a non-terminal point, simply because the opponent was not aware of a move thought to be known to her.

In order to account for these possibilities, for the example presented in Figure 6.3, the 'incomplete' game tree of Figure 6.3b can be seen as a *core game tree* which can be simply extended with duplicates of siblings of all non-leaf nodes in it[1], forming the complete game tree that appears in Figure 6.4. Ideally, the modeller should strategise on the basis of the game tree in Figure 6.4 rather than the one of Figure 6.3b.

6.2.2 The Utility Evaluation Function

The utility value u_i is a value assigned to all nodes in a game tree in two steps. The first is the application of a UEF on each of the leaf-nodes in $\mathcal{GT}$. The second is the propagation of these values upwards by applying the minimax algorithm, with the objective of assigning a utility to the root node N_0, indicating the optimal move for the modeller.

There are multiple ways of defining the utilities of all the possible outcomes —the terminal nodes—of a dialogue game, and all of them are concerned with the modeller's objectives. A naive approach would be to assume that nodes containing

[1]The root node should also be duplicated, though it is ignored in this example, as the dialogue may just consist of the argument in the root node.

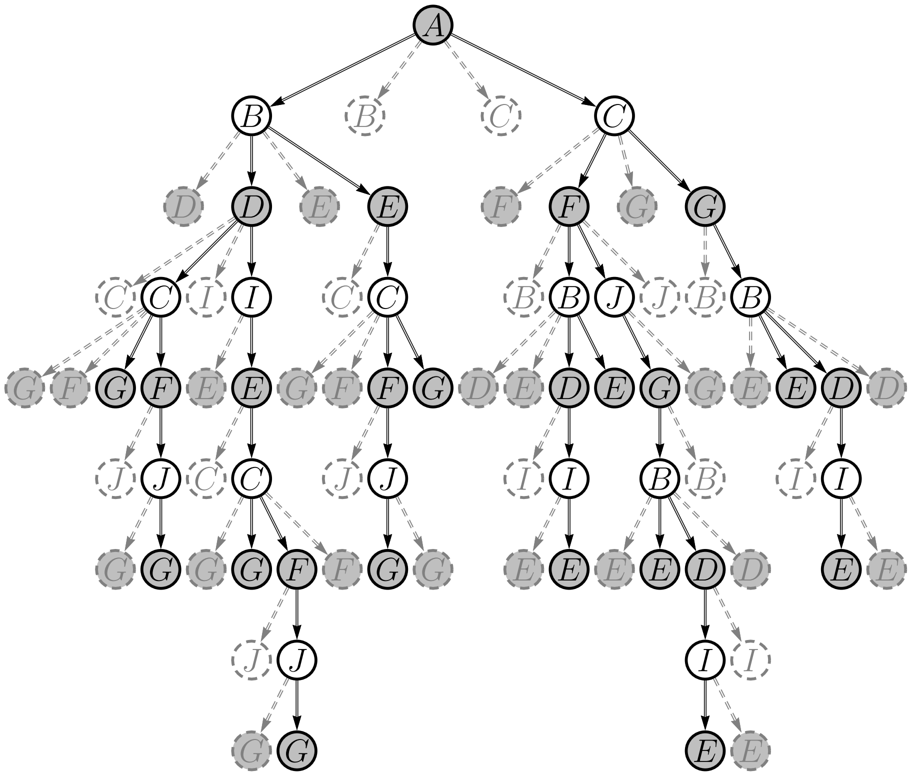
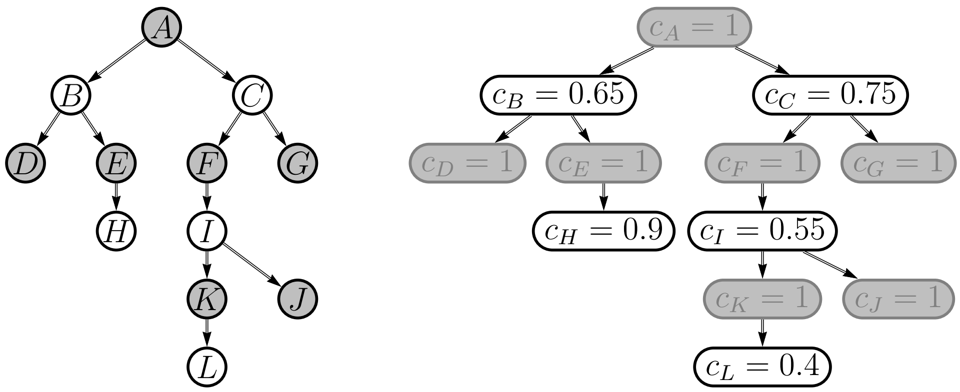


Figure 6.4: The game tree of Figure 6.3b extended to include the duplicate siblings of all non-leaf nodes.



(a) (b)

Figure 6.5: a) A game tree, b) The same game tree with all the arguments replaced with their corresponding confidence values

the same arguments should be assigned the same utility. However, let us again consider this assumption in the case of the example presented in Figure 6.1. In this example, the terminal nodes in both of the possible paths encapsulate the same argument (E). Assuming that both nodes should be assigned the same utility value would disregard the fact that the two nodes are found in different depths in the tree, or that different arguments are revealed by the modeller in each case. Accordingly, if apart from winning it is also essential that this happens as soon as possible, or as late as possible with the objective to stall the opponent, or even if revealing a certain argument should be avoided even if it means losing the game, then the leaf-nodes should receive utility values reflecting such objectives. In general, leaf-nodes should be perceived as discrete outcomes regardless of the arguments they encapsulate.

We provide a formula for assigning such values; one that accounts for the modeller's goals. Since we consider defining a set of goals for the participants in dialogues to be a separate research issue, we only provide a general formula for utility assignation, accounting for the nature of those objectives in an abstract way. However, we are more concrete when it comes to accounting for the confidence values assigned to each of the OAs contained in a certain path leading to a leaf-node, and how these values affect the computation of a utility value.

Assume for example the $\mathcal{GT}$ which appears in Figure 6.5a, where for convenience we assume that it was constructed from a dialogue tree where backtracking is not allowed, while we also ignore the additional terminal states produced if one accounts for imperfect play, and where each node is represented as the argument it encapsulates. In this case the possible outcomes of the dialogue, represented as leaf-nodes in $\mathcal{GT}$, are D, H, L, J & G. Figure 6.5b illustrates the same tree whose nodes have been replaced with the corresponding confidence values of each of the arguments they encapsulate. For the modeller (proponent), reaching any of these leafs depends on two factors. The first is optimally choosing the best move to make when found at a strategic point in a dialogue, which depends on the expected utility values assigned to the possible outcomes of the game. The second is whether the opponent will react as anticipated, i.e. that the modeller's assumptions about its opponent's knowledge are valid and that, given so, the opponent will reasonably act in a predictable way. As the modeller cannot be certain of the validity of the information the modeller holds on its opponent, the modeller can only rely on the confidence values assigned to that information.

For doing so, a leaf utility value should, among others, account for the probability of reaching to a certain leaf, which is equal to the propagation of the confidence values found on the path that leads to the concerned leaf. Since all the proponent arguments are assumed to have a confidence value equal to 1 they are purposely dimmed in the figure as they have no actual effect on how the dialogue will really evolve. Assume for example that we want to calculate the confidence of a path that leads to L, which for convenience we represent as a sequence of arguments, instead of nodes, from A to L in the following form: $P_{A \to L} = \{A, C, F, I, K, L\}$. In this case, assuming a function $\mathtt{C}(P_{A \to L})$ applied on a path in a $\mathcal{GT}$, the modeller's confidence of reaching argument L through it is:

$$\begin{aligned}\mathtt{C}(P_{A \to L}) &= c_A \cdot c_C \cdot c_F \cdot c_I \cdot c_K \cdot c_L \\ &= 1 \cdot 0.75 \cdot 1 \cdot 0.55 \cdot 1 \cdot 0.4 \\ &= 0.165\end{aligned}$$

which is formally defined in Definition 66, bellow:

Definition 66 (Path Confidence) *Let $\mathcal{GT}$ be a game tree and:*

$$P_{0 \to li} =< N_0, \ldots, N_j, \ldots, N_{li} >$$

a path in $\mathcal{GT}$, then $\mathcal{C}$ is a function applied on $P_{0 \to li}$ such that:

$$\mathcal{C}(P_{0 \to li}) = \prod_{j=0}^{li} c_j$$

where c_j is the confidence value found in each node $N_j \in P_{0 \to li}$.

We note that this product assume independence of the component multiplicands, i.e. an argument's confidence value is not dependent on whether arguments that precede it in the path, will or will not appear.

Lastly, let us assume that each leaf-node receives an intermediate utility value calculated according to the modeller's objectives, through the application of some general formula, which we express as u_L (once again, the node index has been replaced with the encapsulated argument for convenience). The final utility value of the concerned leaf-node should be the product of that intermediate utility value (u_L) and $C(P_{A \to L})$. In this respect, we provide the following definition for calculating the utility values of the leaf-nodes of a $\mathcal{GT}$ in the following definition:

Definition 67 (Utility Evaluation Function) *Let $\Pi = \{P_{l1}, P_{l2}, \ldots, P_{lr}\}$ be the set of all possible paths in a $\mathcal{GT}$, and $\mathcal{G} = \{g_1, g_2, \ldots, g_n\}$ be an agent's goals. Then the utility u_{li} of every leaf-node N_{li} in the set $L_N = \{N_{l1}, N_{l2}, \ldots, N_{lr}\}$ is:*

$$u_{li} = U(\mathscr{D}, N_{li}, \mathcal{G}) \cdot \mathcal{C}(P_{li})$$

where U is a function applied on $\mathcal{GT}$:

$$U : \mathscr{D} \times L_N \times \mathcal{G} \to [0, 1]$$

that takes as input the modeller's goals $\mathcal{G}$, the concerned leaf-node, and the dialogue tree $\mathscr{D}$ (from which one can in turn define the $\mathcal{GT}$), and returns a numerical value between 0 and 1.

In essence, function $U_{\mathcal{GT}}$ produces a normalised value in the closed range of $[0, 1]$, in order to guarantee that any possible overall utility value will also be in the same closed range. This is necessary in order for the modeller to be able to evaluate all possible utilities across the same scale. In other words, since both the confidence values as well as the intermediate utility values range between the values of 0 and 1 their product will also exist in the same range, while the fact that these final values are produced based on the same methods makes them comparable.

6.3 Minimax

The next step, after defining the utility values for all the leaf-nodes in a $\mathcal{GT}$, is to also assign similar values to the rest of the nodes in it. This is done by applying the minimax algorithm.

The minimax algorithm is an algorithm applied on a game tree, beginning from its leaf-nodes. The algorithm assumes two roles; those of the *min* and the *max* player respectively representing the opponent and the proponent (the modeller) in the game. Then it traverses the tree towards the root, propagating the leaf utility values upwards selecting interchangeably the minimum or the maximum value to be assigned to the traversed nodes, between the utilities of their successors.

Essentially, the *min* and the *max* roles are respectively assigned to the opponent and the proponent in an attempt to simulate their behaviours in the game. Assuming that both participants are reasonable then we should expect them to try to maximise their respective utilities. Assuming that their utilities are inversely dependent, maximising the proponent's utility implies minimising the opponent's, and vice versa. While when found in strategic points (points where a choice is possible) the proponent is expected to choose those paths which will lead to the highest utility, the opponent is expected to attempt the opposite, i.e. to minimise the proponent's utility and maximise its own by also choosing the appropriate paths when also found in such points in a game.

Let us assume the example depicted in Figure 6.6 which is the game tree (prior to the application of minimax algorithm) for the dialogue tree in Figure 6.5a. No-

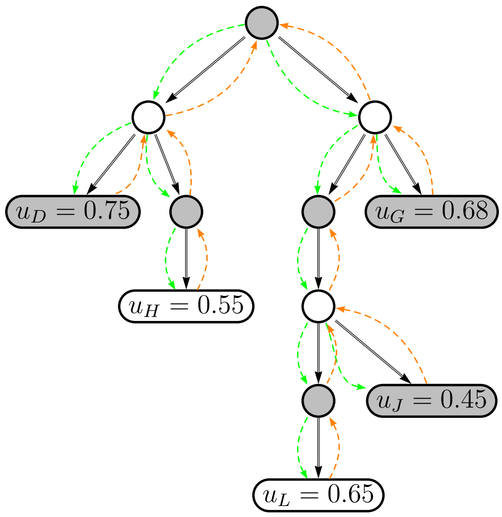


Figure 6.6: The game tree of Figure 6.5a prior to application of the minimax algorithm

tice that all leaf-nodes are assumed to be assigned with an expected utility value after the application of the UEF as described by Definition 67. The algorithm is recursively applied to all the nodes of the tree beginning from the root node (N_0) which in this case is argument A, consistently moving downwards towards the leafs (green dashed arrows arrows). As a convention we assume that the children of each node are traversed from left to right. When a leaf-node is reached the utility value of that node is returned to the father node (orange dashed arrows) and is replaced with the existing value in accordance to whether the existing value is greater or lesser than that of the node's, respectively if the father node is a *min* or a *max* player. Figure 6.7 illustrates how exactly the utility values are propagated upwards, choosing the minimum values when found at opponent nodes and the maximum when found at proponent nodes.

The described algorithm appears in the form of pseudo-code in Algorithm 6.1, and it describes the standard minimax algorithm found in many text books such as Hazewinkel [2001]. We note that the initial call values for both N_i and the `maxPlayer` (which we assume to be a boolean variable) are:

$$minimax(N_0, FALSE)$$

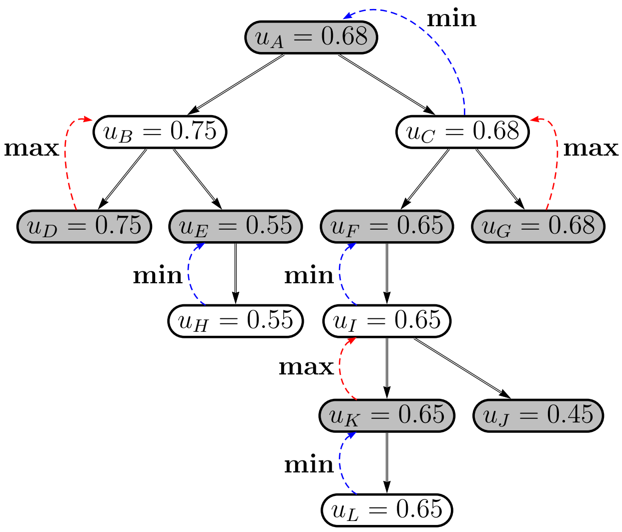


Figure 6.7: The same game tree after the application of the minimax algorithm, with all the arguments replaced with their corresponding confidence values

Once all nodes of the game tree have been assigned a utility value, the optimal decision making problem is resolved by choosing, when found at a certain father-node, the path formed by the child-node with the highest utility. In the case of the example in Figure 6.5a, the modeller (the proponent) is found at a decision making point at node C (after the opponent opts to move argument C over B against the root node A). At node C the optimal choice for the modeller appears to be argument G with a utility value $u_G = 0.68$.

6.4 Conclusions

In this chapter we have provided a strategising mechanism which utilises the logical information collected using the different collection methods proposed in Chapter 4. Particularly, we illustrated how the confidence values associated with the logical elements of OM can be used to affect the way that a UEF assigns values to the terminal states of a game.

Since applying a UEF on a dialogue tree is not possible, we define a corresponding game tree; a problem similar to finding all the possible, same source, discrete paths to multiple destinations in a graph. Essentially, having provided

```
ALGORITHM minimax(N_i, maxPlayer)
if N_i ∈ L_N then
    return u_i
end if
if maxPlayer then
    bestValue := 0
    for all childs of N_i do
        val := minimax(child, false)
        bestValue := max(bestValue, val)
        return bestValue
    end for
else
    bestValue := 1
    for all childs of N_i do
        val := minimax(child, true)
        bestValue := min(bestValue, val)
        return bestValue
    end for
end if
```

Algorithm : 6.1: Minimax algorithm

a definition concerned with the instantiation of a dialogue tree's reply structure, which encapsulates all of the replies in a dialogue tree in the form of pairs of arguments, and given both a dialogue tree and a game tree, then according to Definition 65, we say that the game tree is the dialogue tree's corresponding game tree if for every element in a game tree's path, there exist a preceding element, where the pair of arguments encapsulated in both of them—ordered with the argument in the preceding node appearing second—are members of the dialogue tree's reply structure.

We then illustrated how a general version of UEF can account for the confidence values of the OAs found on the paths that lead to an evaluated leaf-node, by multiplying their propagation product with the leaf's utility estimate, and we finally showed how the minimax algorithm can be applied on the utilities of a game tree's leaf-nodes, in order to propagate them upwards towards the root showing how decision making at each point of a dialogue can then be defined by the utility values of the possible choices.

As in this thesis we do not formally define the possible goals that participants may have in dialogues, testing of the effectiveness of the minimax algorithm application remains an open issue. In addition, accounting for counter-strategising is also an issue worth investigating, but what matters more is evaluating the validity of the collected information, on which an agent's strategising relies. However, the development of benchmarks for argument-based dialogue platforms is only recently gaining interest (Cerutti et al. [2014]). Thus, in the next chapter we focus on proposing a methodology towards evaluating the validity and the effectiveness of the information collection methods proposed in this thesis, and leave evaluation of the strategising processes to future work.

Chapter 7

Towards a Methodology for Evaluation

The importance of developing mechanisms for building and updating opponent models (OMs) has been repeatedly stressed throughout this thesis. An OM is an essential ingredient in strategising, and thus success of any strategising approach which relies on OMs largely depends on their validity. The formal mechanisms defined in Chapters 4 and 5, which relate opponent information and allow for the augmentation of OMs, rely on the interrelatedness of information expressed through two notions; those of *support* and *common attack targets*. Modelling opponent knowledge based on these notions needs to be tested with respect to whether they can actually achieve an increase in an OM's credibility. However, due to the lack of argument-based benchmarks for agent dialogues, both approaches remain untested.

In this chapter we discuss and propose a methodology for evaluating the success of the proposed modelling approaches in producing valid models of opponent information, particularly concerned with the augmentation process. In addition, we describe an extended example which summarises our work in augmenting OMs in order to illustrate how one could evaluate our approach, given an environment able to account for the logical representation of agent knowledge at a structural level, the construction of arguments, and the scalability and complexity of multiple agent interactions.

7.1 Introduction

In Chapters 4 and 5 we defined a process that allows the construction and expansion of OMs, which relies on the interrelatedness of information. Development of this approach builds on the assumption that inherent relationships between arguments, such as that of support or that of arguments which share common attack targets, can be quantified, through monitoring patterns of their appearances in an agent's general dialogue history, so as to allow the modeller to anticipate what might follow in a dialogue against a certain opponent. Nevertheless, the extent to which this quantification captures the actual strength of these relationships, leading to a useful incorporation of additional information to an OM, remains untested.

What has been tested and evaluated so far is the tractability of our approach, in Chapter 4, for the purpose of which we relied on Poisson random graphs. For the production of these graphs we ensured accounting for simple graph properties that we expect argument graphs to be characterised with, such as that they will not be complete or have a path-like or grid-like structure. Accounting for additional structural properties of argument graphs requires a thorough analysis of numerous factors mostly concerned with the employed argumentation logics, and this is what was attempted by Hunter and Woltran [2013]. In their work, they researched a number of logical argument systems and associated them with certain classes of argument graphs they could induce, such us rooted, acyclic, weakly connected, self-loop, or complete graphs. However, they note that issues raised in their work may have ramifications particularly in relation to systems such as $ASPIC^+$ (Prakken [2010]) and ABA (Dung et al. [2009]).

Relying on random graphs for testing the tractability of our approach is adequate for evaluating how quickly we can produce good estimates of an argument's relationship likelihood to an OM. It is inadequate though for deciding on whether augmenting an OM will actually be effective in increasing its validity. Such testing would have to rely on actual argument relationships graphs ($\mathcal{R}G$s), since even if one could induce random argument graphs that would satisfy, in our case, all the structural properties of an $ASPIC^+$ induced graph, it is still questionable whether the produced results would be valid. This is due to the fact that the interrelatedness of information is more likely to be captured through the incremental construction of a $\mathcal{R}G$, built through a series of agent dialogues. In contrast, a randomly generated graph may ignore this feature, as arguments will be randomly linked irrespective of their relationship in dialogues. In addition, since quantification of these relationships will also be random, it will fail to capture the statistical information on which one should rely. Thus, any conclusions on the effectiveness of our approach that would rely on randomly generated graphs would at least be questionable.

Furthermore, evaluation of the impact of accounting for the underlying logic in dialogues and subsequently for the possible construction of new arguments within dialogues (i.e. accounting for the logical instantiation of arguments which is one of the key contribution of this thesis stressed in Chapter 3), is also an issue that needs to be evaluated. In other words, it is important to test through a case study the possible outcomes of numerous dialogues, comparing cases where the underlying logic *is* and *is not* accounted for. This way we could investigate the chain of events that might evolve, as well as the consequences that result, from the instantiation of a new argument, either in a single dialogue or in a series of dialogues, as well as in an agent's knowledge base.

Unfortunately, practical evaluation of argumentation research through the use of appropriate general purpose mechanisms for benchmarking, is only now gaining interest (Bex et al. [2014]; Cerutti et al. [2014]; Thimm [2014]). Interest in this research topic is initially recognised in the work of Bistarelli et al. [2013], which is a first attempt to compare three different implementations of AAS. For the purpose of their work they also resort to randomly generated test networks.

In an effort to address this issue Cerutti et al. [2014] attempt the development of a benchmark framework named `Probo`, which aspires to serve as a platform for easily comparing different implementations for solving argumentation problems. It however focusses only on abstract argumentation, though extensions to structured argumentation have been mentioned in their future work. In contrast, the work of Thimm [2014], offers a more flexible, and widely applicable, knowledge representation framework, referred to as `Tweety`, in the form of a set of `Java`-based libraries which, among others, allows for the representation of structured argumentation. Interestingly, dialogue libraries are also offered, while they report on two case studies of strategic argumentation in an attempt to illustrate how `Tweety` may be used for empirical evaluation purposes.

Lack of general purpose mechanisms in the context of dialogue games, able to handle high level executable definitions concerned with classical features of argument dialogue games, such as: the handling of commitment stores; rules for turn-taking, or; the dialogical construction of underlying argument structures; is also reported in the work of Bex et al. [2014]. In their work, they rely on certain languages for the specification of dialogue protocols in a machine-readable way, allowing for the protocol specifications to be expressed separately from the program that executes them. Assuming extension of these languages aimed specifically at argumentation dialogues, they develop a Dialogue Game Execution Platform (DGEP) which, as they explain, is not only able to execute these dialogue protocol specifications, but can also handle connections to software agents, argument structures, large knowledge bases and human-computer interfaces. Among others, DGEP is capable of producing schematic representations of a single move in a dialogue, its reply and the connections to the underlying argument structure in terms of the argument interchange format (AIF) ontology (Chesñevar et al. [2006]).

Finally, an $ASPIC^{+}$ `Java`-based argument evaluation platform has been recently developed by Snaith and Reed. [2012], referred to as `TOAST`. Provided a set of premises and a set of rules with associated preference and contrariness information, `TOAST` can produce visual information on the acceptability of arguments in the derived argumentation framework. As such, it is mostly used for medical reasoning purposes, while future extensions of the platform through the inclusion of appropriate libraries for the possible accommodation of dialogue frameworks is not discussed.

In our case, the need of empirically evaluating our approach through a case study is evident, which makes the use of such platforms imperative. Nonetheless, since their establishment is only now taking place, we simply present a methodology towards evaluation, through an extended example which summarises our work on the augmentation of OMs, and leave its application to future work. We note, again, that evaluating the impact of accounting for the dynamic instantiation of arguments by accounting for the underlying logic that characterises the exchanged arguments in a dialogue, is an important issue that needs to be evaluated as well, and this is also something we are interested in researching in future work.

The rest of the chapter is structured as follows: in Section 7.2 we present two evaluation metrics used to characterise the validity of OMs; in Section 7.3 we present a detailed example, illustrating: the incremental construction of an agent's $\mathcal{R}G$; an augmentation of a certain OM, and; an evaluation of the augmented OM with the corresponding actual knowledge of that opponent is performed, relying on the proposed evaluation metrics. Lastly we discuss another, more strategically oriented, way of evaluating our augmentation approach in Section 7.4, and summarise our work in Section 7.5.

7.2 Evaluation Metrics

Throughout this thesis an OM is represented as a sub-theory $S_{(i,j)} = \langle AT_{(i,j)}, \mathcal{G}_{(i,j)} \rangle$, where $AT_{(i,j)}$ is what Ag_i believes is the argumentation theory (Definition 17) of Ag_j and $\mathcal{G}_{(i,j)}$ is what Ag_i believes are the goals of Ag_j. For convenience, a sub-theory is informally expressed as a tuple of sets of structural logical elements $S_{(i,j)} =< \mathcal{K}_{(i,j)}, \leq'_{(i,j)}, \mathcal{R}_{(i,j)}, \leq_{(i,j)}, \mathcal{G}_{(i,j)} >$, where, $\mathcal{K}$ is the set of premises, $\mathcal{R}$ the set of rules, $\leq'$ and $\leq$ their respective preference-orderings, and G the set of goals, that Ag_i assumes Ag_j to be aware of.

Evaluation of an OM essentially concerns the level at which the collected information ($S_{(i,j)}$) about a certain opponent accurately reflects that opponent's actual knowledge ($S_{(j,j)}$). The extent to which there exists a correspondence between these two sets is a measure of the validity and thus credibility of an OM. Specifically, this correspondence can be distinguished in two ways: one concerns how *correct* an OM is, and; the other concerns how *complete* it is. Essentially, the *correctness* of an OM relates to the question: *How much of the modelled information is contained in the opponent's knowledge?* Similarly, the completeness of an OM relates to the question: *How much of the opponent's knowledge is contained in the modelled information?*

These two correspondence types are formally defined as follows:

Definition 68 (Correct Opponent Model) *Let $S_{(i,j)}$ be Ag_i's OM of Ag_j, and $S_{(j,j)}$ be Ag_j's actual knowledge. Then we say that $S_{(i,j)}$ is* `Correct` *iff:*

$$\alpha \in S_{(i,j)} \longrightarrow \alpha \in S_{(j,j)}$$

$\forall \alpha \in S_{(i,j)}$.

Definition 69 (Complete Opponent Model) *Let $S_{(i,j)}$ be Ag_i's OM of Ag_j, and $S_{(j,j)}$ be Ag_j's actual knowledge. Then we say that $S_{(i,j)}$ is* `Complete` *iff:*

$$\alpha \in S_{(j,j)} \longrightarrow \alpha \in S_{(i,j)}$$

$\forall \alpha \in S_{(j,j)}$.

We note that strictly speaking one can never know with absolute certainty an opponent's actual knowledge, unless of course this is provided within the context of a controlled experiment.

Given these definitions, the validity of an OM can be decided with respect to whether the model is or is not `Complete` as well as `Correct`. However, different levels of soundness and completeness can be further distinguished, in order to provide a more fine grained representation of an OM's correspondence to an opponent's actual knowledge. The typifications introduced in Chapter 3, Section 3.1.1, serve exactly for this purpose. To recap we assume that an OM $S_{(i,j)}$ is:

- `False`: iff $S_{(i,j)} \cap S_{(j,j)} = \varnothing$
- `Partial`: iff $S_{(i,j)} \cap S_{(j,j)} \neq \varnothing$
- `Contained`: iff $S_{(i,j)} \subset S_{(j,j)}$
- `Identical`: iff $S_{(i,j)} = S_{(j,j)}$
- `Excessive`: iff $S_{(i,j)} \supset S_{(j,j)}$

Relying on these typifications, as well as on the notions of correctness and completeness, one may characterise an OM as not `Correct` and `Partial`, or not

`Complete` and `Contained`, or even both `Correct` & `Complete` and thus `Identical`. Furthermore, if an even more precise representation of the correspondence between the two sets is desired, then these typifications may also be used to characterise the level of correspondence of each of the assumed sets of structural logical elements $\mathcal{K}, \leq', \mathcal{R}, \leq, \mathcal{G}$ of a sub-theory $S_{(i,j)}$ to $S_{(j,j)}$, e.g. the correspondence of $\mathcal{K}_{(i,j)}$ to $\mathcal{K}_{(j,j)}$.

Given these, a first approach to evaluation is to investigate the extent to which any of these characterisations of an OM is possible within our framework. In the context of this thesis, provided that we rely on a defeasible reasoning model which implies that information is never discarded, then, without loss of generality, it is safe to assume that the accumulated information specifically concerned with the premises and inference rules asserted by a particular opponent in a dialogue game, will always be part of that opponent's knowledge.

In this respect, let S^- represent an OM from which priority orderings as well as goals are excluded, then the following proposition holds:

Proposition 4 *Let $S^-_{(i,j)}$ be Ag_i's OM of Ag_j's actual knowledge $S^-_{(j,j)}$, built* solely *from Ag_i's direct collection of information in dialogues against Ag_i. Then:*

$$S^-_{(i,j)} \subseteq S^-_{(j,j)}$$

Proof Straightforward, given that, whatever is acquired through a dialogue process and based on direct collection will be added to the OM, while it will never be discarded from the opponent's knowledge.

In other words, information about a particular agent, collected *solely* through direct collection particularly related with the rules ($\mathcal{R}$) and the premises ($\mathcal{K}$) introduced by that opponent, can neither be `False` nor `Excessive`. However, if one additionally accounts for third party information, and augmentation information, then $S^-_{(i,j)} \subseteq S^-_{(j,j)}$ can no longer be guaranteed, while the creation of either a `False`, or an `Excessive` OM becomes possible.

Hence, when faced with a strategic choice of move, an agent should favour decisions based on information directly collected by itself over information provided by third party agents or which resulted from an augmentation process. Of course, this is implicitly already modelled in the proposed approach, given that all information collected by a participant is assigned a credibility level equal to 1, as opposed to third party information which is assigned a credibility value based on the result of the trust function described in Definition 46, and information added to the model through an augmentation which is associated with a likelihood value.

In overall, differentiating between information in an OM based on its origin can result in increasing the credibility of the model with respect to its *correctness*, and thus consequently to increase the effectiveness of the strategy function that relies on it. Additionally, a participant has to also account for the possibility that she may be unaware of additional information possibly known to her opponent. In this respect, an OM can be characterised as credible if it is to a large extent *correct*, and; a good approximation of all its knowledge, i.e. as *complete* as possible. Practically, assuming two sets $S_{(i,j)}$ and $S_{(j,j)}$, the correctness of $S_{(i,j)}$ can be computed by dividing the number of elements in the intersection of the two sets by the number of elements in $S_{(i,j)}$. Obviously, the denominator in this case will never have a higher value than the enumerator, thus always producing a rational value in the interval $[0, 1]$. Similarly, the completeness of $S_{(i,j)}$ is computed, again, by dividing the number of elements in the intersection of the two sets, but this time, by the number of elements in $S_{(j,j)}$.

Relying on these two metrics we can decide on the correspondence *accuracy* of an OM, which can be relatively expressed as the product of its correctness and completeness. Basically, given multiplication, the more correct and complete then the more accurate an OM will be. One may then, compare the overall accuracy of two models by comparing the products of their corresponding correctness and completeness. Though these metrics can be applied on any of the corresponding logical sets of the two compared sub-theories, we present, in Definition 70, a high level accuracy metric that only concerns the relationship between the sets of arguments instantiated from them.

Definition 70 (Accuracy) *Let S be an agent's sub-theory and $\mathcal{A}$ the set of arguments instantiated from S. Then, assuming two sub-theories $S_{(i,j)}$ & $S_{(j,j)}$ where $\mathcal{A}_{(i,j)}$ & $\mathcal{A}_{(j,j)}$ are the respective sets of arguments instantiated from them,* *then the* **`correctness`** *of* $S_{(i,j)}$ *referred to as* $\varrho_{(i,j)} \in [0,1]$*, is:*

$$\varrho_{(i,j)} = \frac{|\mathcal{A}_{(i,j)} \cap \mathcal{A}_{(j,j)}|}{|\mathcal{A}_{(i,j)}|}$$

the **`completeness`** *of* $S_{(i,j)}$*, referred to as* $\omega_{(i,j)} \in [0,1]$*, is:*

$$\omega_{(i,j)} = \frac{|\mathcal{A}_{(i,j)} \cap \mathcal{A}_{(j,j)}|}{|\mathcal{A}_{(j,j)}|}$$

while the **`accuracy`** *of* $S_{(i,j)}$*, referred to as* $\hat{\alpha}_{(i,j)} \in [0,1]$*, is:*

$$\hat{\alpha}_{(i,j)} = \varrho_{(i,j)} \cdot \omega_{(i,j)}$$

In a trivial example, assuming two sub-theories $S_{(1,2)}$ and $S_{(2,2)}$ and the corresponding sets of arguments instantiated from them $\mathcal{A}_{(1,2)}$ and $\mathcal{A}_{(2,2)}$ where:

$$\mathcal{A}_{(1,2)} = \{A, B, C, D, E, X\}$$

and

$$\mathcal{A}_{(2,2)} = \{A, B, C, D, E, W, Y\}$$

then the `correctness` of $S_{(1,2)}$ would be:

$$\varrho_{(1,2)} = \frac{|\mathcal{A}_{(1,2)} \cap \mathcal{A}_{(2,2)}|}{|\mathcal{A}_{(1,2)}|} = \frac{5}{6}$$

its `completeness`:

$$\omega_{(1,2)} = \frac{|\mathcal{A}_{(1,2)} \cap \mathcal{A}_{(2,2)}|}{|\mathcal{A}_{(2,2)}|} = \frac{5}{7}$$

while its `accuracy` to $S_{(2,2)}$, would be:

$$\hat{\alpha}_{(1,2)} = \varrho_{(1,2)} \cdot \omega_{(1,2)} = \frac{5}{6} \cdot \frac{5}{7} = \frac{25}{42} \approx 0.6$$

7.3 An Extended Evaluation Example

In this section we present, in a series of detailed steps, an extended example in an attempt to illustrate how one could evaluate our approach in an environment able to account for the scalability and complexity of multiple agent dialogue interactions. We begin by stating a number of necessary assumptions.

1. All agents share the same contrary relation, the same language $\mathcal{L}$, and the same way of defining preferences over arguments, i.e. all agents share the same function p (Definition 17).

2. The incrementally constructed $\mathcal{RG}$ will be built for a θ_t threshold equal to 1 and in accordance to Definition 51;

3. Weighting on the arcs of the constructed $\mathcal{RG}$ will follow the Conductivity-based Weight Assignation definition (Definition 60);

4. Agents may engage in both grounded as well as credulous dialogue games, where respectively the proponent (Pr) cannot repeat the same move in a dispute while the opponent (Op) can, and vice versa;

5. The instances where an opponent argument (OA) is repeated in a dispute will be ignored, as they can only be interpreted as self-supporting relationships and thus have no impact on our augmentation approach (refer to discussion in Section 5.2.1);

6. With respect to the knowledge of the participating agents, rather than looking at the sub-theories that each agent holds with respect to one's own ($S_{(i,i)}$) or one's opponent's knowledge ($S_{(i,j)}$), we are basically representing these sub-theories as argument graphs. We will therefore be representing an $S_{(i,j)}$ as $\mathcal{A}_{(i,j)}$, which represents the set of arguments instantiated from $S_{(i,j)}$, and;

7. Finally, though instantiation of additional arguments by an agent is generally possible through the incorporation of structural information, by means of any of the three ICMs, into an OM, since this is an abstract example

we will assume that no such arguments appear throughout the evaluation process.

In addition to these assumptions, it is important to remind the reader that in our framework we do not account for how agents update their knowledge based on arguments submitted in dialogues by their opponents (Section 3.5). Hence, in the dialogues that follow in this chapter, while an agent Ag_i may use some argument X which was previously introduced in other dialogues by $Ag_{j \neq i}$ (other dialogues between Ag_i and $Ag_{j \neq i}$) introduction of X by Ag_i will not be because Ag_i has updated his knowledge by including argument X in it, but because that argument was already known to Ag_i.

Furthermore, we note that for evaluation purposes it is necessary that complete knowledge of all the information used in the dialogues that will appear is provided prior to the commencement of those dialogues. In other words, we assume the role of an observer from the outset to whom the distinct knowledge of every dialogue participant is known. We refer to the sum of all these knowledge bases as Multi-Agent Omni-Base (MAOB) (refer to Table3.1a), and we represent the set of arguments instantiated from them as Omni-Graph ($O\mathcal{G}$).

Example 11 *Let us assume a set of agents* $Ags = \{Ag_1, Ag_2, Ag_3, Ag_4, Ag_5, Ag_6\}$*, where* Ag_1 *is the modeller who will be interacting with the rest of the agents through a series of 10 dialogues. Also, let the knowledge of* Ag_6*, be composed of the following arguments:*

$$\mathcal{A}_{(6,6)} = \{W, Y, T, U, E, O, A, F, P, C, H, V, L, J, Z\}$$

and the complete knowledge known by all agents be expressed as an $O\mathcal{G}$*, whose form appears in Figure 7.1. The evaluation process is broken down to a series of steps as follows:*

1. *A series of 10 dialogues will be presented, between* Ag_1 *and agents* Ag_2, Ag_3*,* Ag_4 *and* Ag_5 *along with the incremental construction of a* $\mathcal{R}G_1$ *by the modeller (*Ag_1*). Edge-weights will be concurrently computed.*

2. *After construction of the* $\mathcal{R}G_1$*, agent* Ag_1 *will engage in another dialogue with* Ag_6*.*

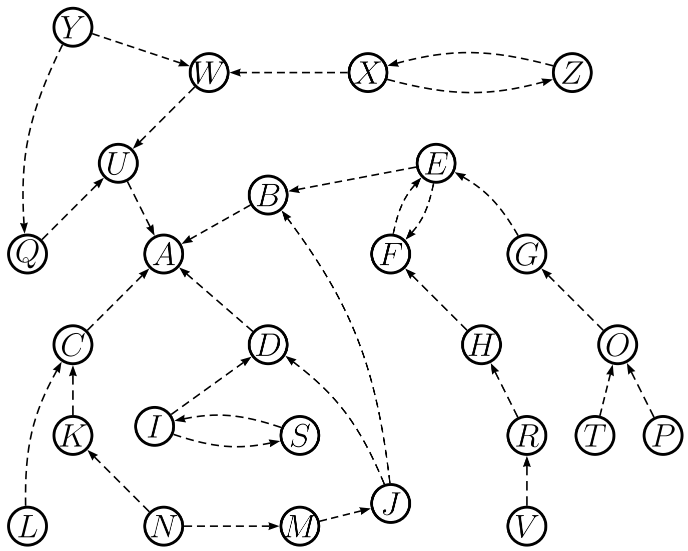


Figure 7.1: The omni-graph which comprises the complete knowledge of all agents participating in the dialogues.

3. *At the end of this latter dialogue, Ag_1 will produce a model of Ag_6's knowledge ($\mathcal{A}_{(1,6)}$), solely based on direct collection of information from commitment stores.*

4. *After construction of $\mathcal{A}_{(1,6)}$ the modeller, Ag_1, relying on $\mathcal{RG}_1$ will attempt an* off-line augmentation *of it:*

$$f_{aug}(\mathcal{A}_{(1,6)}) \mapsto \mathcal{A}'_{(1,6)}$$

in order to incorporate additional arguments based on the augmentation process described in Chapter 4;

5. *Finally, after the augmentation takes place, assuming knowledge of the actual information known to Ag_6, expressed as $\mathcal{A}_{(6,6)}$, the* **`accuracy`** *between the augmented OM ($\mathcal{A}'_{(1,6)}$) and $\mathcal{A}_{(6,6)}$, as well as between the non-augmented OM ($\mathcal{A}_{(1,6)}$) and $\mathcal{A}_{(6,6)}$ will be computed and compared.*

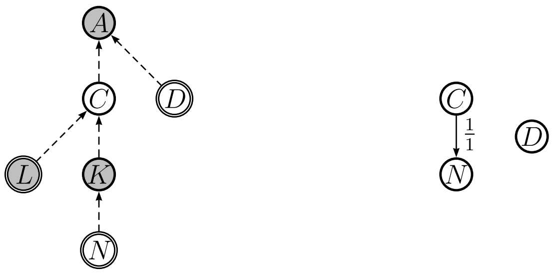
We note that reference to a dialogue move will be expressed as a tuple $< I, \mathcal{DM}_i^{con} >$ where $I \in \{Pr, Op\}$ with Pr and Op being replaced with the corresponding agents participating in a dialogue, and $\mathcal{DM}_i^{con}$ is the argument in the corresponding dialogue move, which is a slightly different representation of the form provided in Definition 20. Also, for presentation convenience a dialogue sequence (Definition 21) will be expressed as a sequence of such tuples, e.g.:

$$\mathscr{D}_1 =<< Ag_1, A >, < Ag_2, B >, \ldots, < Ag_1, F >>$$

Additionally, in all dialogues, and in contrast to the colouring convention maintained throughout the thesis so far according to which the proponent's moves appear in grey colour and the opponent's moves in white, in this example the modeller's moves will appear in grey colour as opposed to all other moves which appear in white.

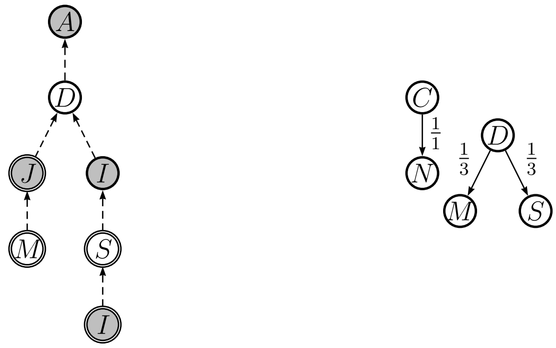
***Step** 1: **Incremental Construction of** $\mathcal{R}G_1$*

***Dialogue 1:** Ag_1 vs Ag_2 on the grounded acceptability of A. Arguments are moved in the following order: $\mathscr{D}_1 =<< Ag_1, A >, < Ag_2, C >$*



(a) (b)

Figure 7.2: a) Dialogue $\mathscr{D}_1$, b) the corresponding $\mathcal{RG}_1$ for Ag_1

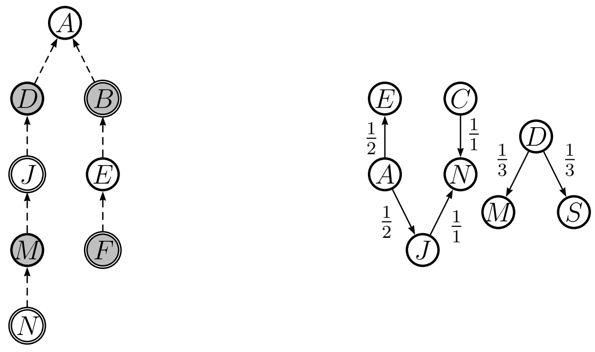


(a) (b)

Figure 7.3: a) Dialogue $\mathscr{D}_2$, b) the correspondingly extended $\mathcal{RG}_1$

$, < Ag_1, K >, < Ag_2, N >< Ag_1, L >, < Ag_2, D >>$. Winner of the dialogue is Ag_2 whose last move (D) results in A being labelled ***`out`****. The dialogue along with the correspondingly constructed $\mathcal{RG}_1$ is shown in Figure 7.2.*

Dialogue 2: *Ag_1 vs Ag_3, on the credulous acceptability of A. Arguments are moved in the following order: $\mathscr{D}_2 =<< Ag_1, A >, < Ag_3, D >, < Ag_1, J >, < Ag_3, M >, < Ag_1, I >, < Ag_3, S >, < Ag_1, I >>$. Winner of the dialogue is Ag_1 whose last move (I which is repeated in the game) results in A being labelled* ***`in`****. The dialogue along with the correspondingly extended $\mathcal{RG}_1$ is shown in Figure 7.3.*



(a) (b)

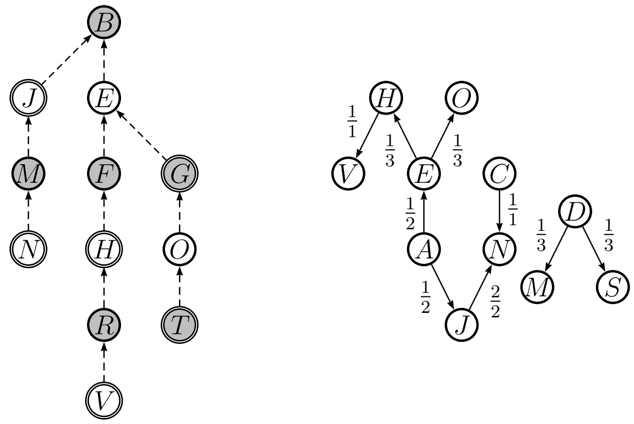
Figure 7.4: a) Dialogue $\mathscr{D}_3$, b) the correspondingly extended $\mathcal{R}G_1$

Dialogue 3: *Ag_4 vs Ag_1, on the grounded acceptability of A. Arguments are moved in the following order: $\mathscr{D}_3 =<< Ag_4, A >, < Ag_1, D >, < Ag_4, J >, < Ag_1, M >, < Ag_4, N >, < Ag_1, B >, < Ag_4, E >, < Ag_1, F >>$. Winner of the dialogue is Ag_1 whose last move (F) results in A being labelled* ***`out`***. *The dialogue along with the correspondingly extended $\mathcal{R}G_1$ is shown in Figure 7.4.*

In relation to the concurrently instantiated weight values, notice for example that specifically w_{DM} and w_{DS} have values equal to $\frac{1}{3}$ and not $\frac{1}{2}$. *This is because even though argument D has appeared in 2 dialogues so far, it participates in 3 disputes within those dialogues: once in $\mathscr{D}_1$ in which it is followed by nothing (i.e. it is followed by the* ***`null`*** *argument), and; twice in $\mathscr{D}_1$ followed respectively by arguments M and S.*

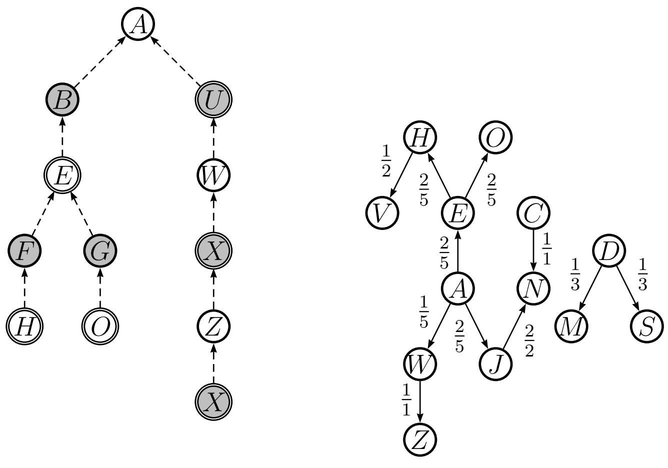
Dialogue 4: *Ag_1 vs Ag_5, on the grounded acceptability of B. Arguments are moved in the following order: $\mathscr{D}_4 =<< Ag_1, B >, < Ag_5, E >, < Ag_1, F >, < Ag_5, H >, < Ag_1, R >, < Ag_5, V >, < Ag_1, G >, < Ag_5, O >, < Ag_1, T >, < Ag_5, J >, < Ag_1, M >, < Ag_5, N >>$. Winner of the dialogue is Ag_5 whose last move (N) results in B being labelled* ***`out`***. *The dialogue along with the correspondingly extended $\mathcal{R}G_1$ is shown in Figure 7.5.*

Dialogue 5: *Ag_2 vs Ag_1, on the grounded acceptability of A. Argu-*



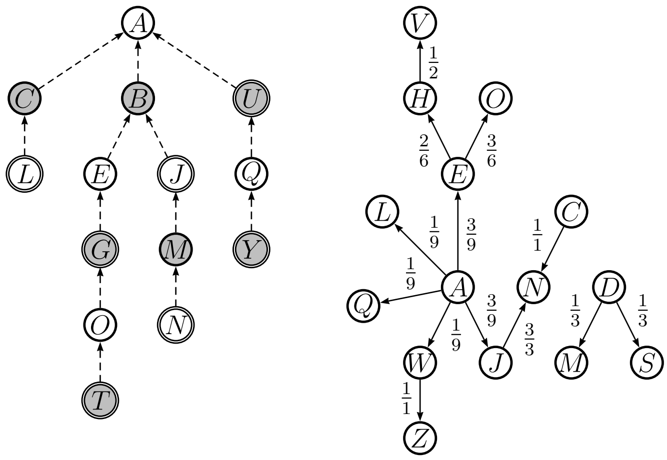
(a) (b)

Figure 7.5: a) Dialogue $\mathscr{D}_4$, b) the correspondingly extended $\mathcal{R}G_1$



(a) (b)

Figure 7.6: a) Dialogue $\mathscr{D}_5$, b) the correspondingly extended $\mathcal{R}G_1$



(a) (b)

Figure 7.7: a) Dialogue $\mathscr{D}_6$, b) the correspondingly extended $\mathcal{R}G_1$

ments are moved in the following order: $\mathscr{D}_5 =<< Ag_2, A >, < Ag_1, B >, < Ag_2, E >, < Ag_1, F >, < Ag_2, H >, < Ag_1, G >, < Ag_2, O >, < Ag_1, U >, < Ag_2, W >, < Ag_1, X >, < Ag_2, Z >, < Ag_1, X >>$*. Winner of the dialogue is* Ag_1 *whose last move (*X *which is repeated) results in* A *being labelled* **`out`**. *The dialogue along with the correspondingly extended* $\mathcal{R}G_1$ *is shown in Figure 7.6.*

Dialogue 6: Ag_3 *vs* Ag_1*, on the grounded acceptability of* A*. Arguments are moved in the following order:* $\mathscr{D}_6 =<< Ag_3, A >, < Ag_1, C >, < Ag_3, L >, < Ag_1, B >, < Ag_3, E >, < Ag_1, G >, < Ag_3, O >, < Ag_1, T >, < Ag_3, J >, < Ag_1, M >, < Ag_3, N >, < Ag_1, U >, < Ag_3, Q >, < Ag_3, Y >>$*. Winner of the dialogue is* Ag_1 *whose last move (*Y*) results in* A *being labelled* **`out`**. *The dialogue along with the correspondingly extended* $\mathcal{R}G_1$ *is shown in Figure 7.7.*

Dialogue 7: Ag_4 *vs* Ag_1*, on the grounded acceptability of* A*. Arguments are moved in the following order:* $\mathscr{D}_7 =<< Ag_4, A >, < Ag_1, B >, < Ag_4, E >, < Ag_1, F >, < Ag_4, H >, < Ag_1, R >, < Ag_4, V >, < Ag_1, G >, < Ag_4, O >, < Ag_1, P >, < Ag_4, J >, < Ag_1, M >, < Ag_4, N >, < Ag_1, C >, < Ag_4, K >, < Ag_1, N >>$*. Winner of the dialogue is* Ag_1

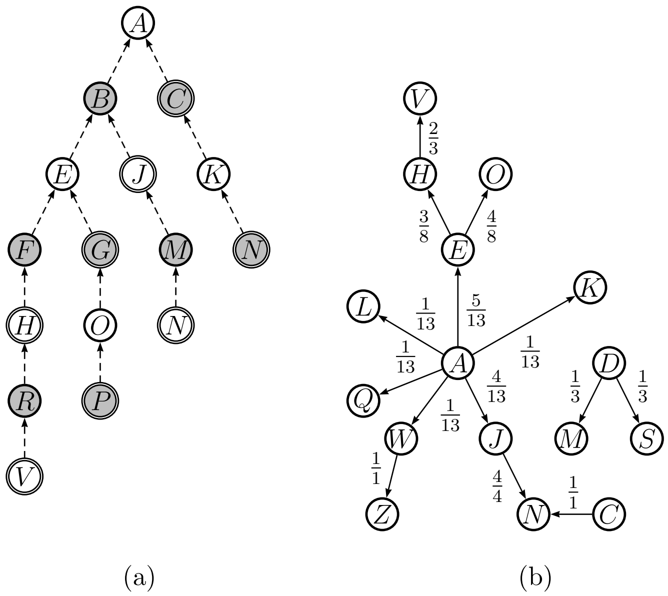


Figure 7.8: a) Dialogue $\mathscr{D}_7$, b) the correspondingly extended $\mathcal{R}G_1$

whose last move (N) results in A being labelled **`out`**. *The dialogue along with the correspondingly extended $\mathcal{R}G_1$ is shown in Figure 7.8.*

Dialogue 8: *Ag_1 vs Ag_2, on the grounded acceptability of A. Arguments are moved in the following order: $\mathscr{D}_8 =<< Ag_1, B >, < Ag_2, E >, < Ag_1, F >, < Ag_2, H >, < Ag_1, R >, < Ag_2, V >, < Ag_1, G >, < Ag_2, O >, < Ag_1, T >>$. Winner of the dialogue is Ag_1 whose last move (T) results in B being labelled* **`out`**. *The dialogue along with the correspondingly extended $\mathcal{R}G_1$ is shown in Figure 7.9.*

Dialogue 9: *Ag_1 vs Ag_4, on the credulous acceptability of U. Arguments are moved in the following order: $\mathscr{D}_9 =<< Ag_1, U >, < Ag_4, W >, < Ag_1, X >, < Ag_4, Z >, < Ag_1, X >, < Ag_4, Q >, < Ag_1, Y >>$. Winner of the dialogue is Ag_1 whose last move (Y) results in U being labelled* **`in`**. *The dialogue along with the correspondingly extended $\mathcal{R}G_1$ is shown in Figure 7.10.*

Dialogue 10: *Ag_1 vs Ag_3, on the grounded acceptability of E. Arguments are moved in the following order: $\mathscr{D}_{10} =<< Ag_3, E >, < Ag_1, F >$*

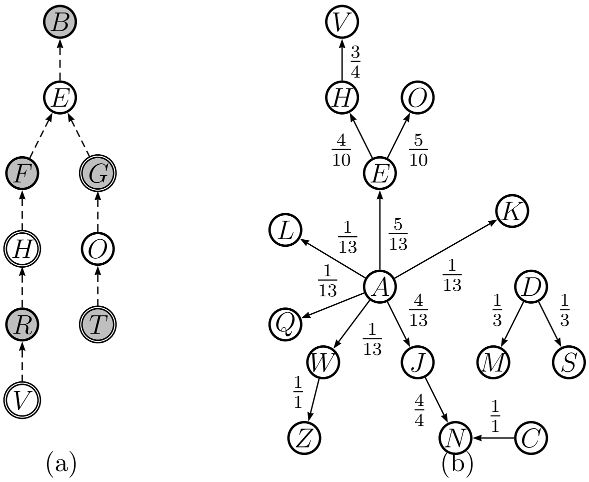


Figure 7.9: a) Dialogue $\mathscr{D}_8$, b) the correspondingly extended $\mathcal{R}G_1$

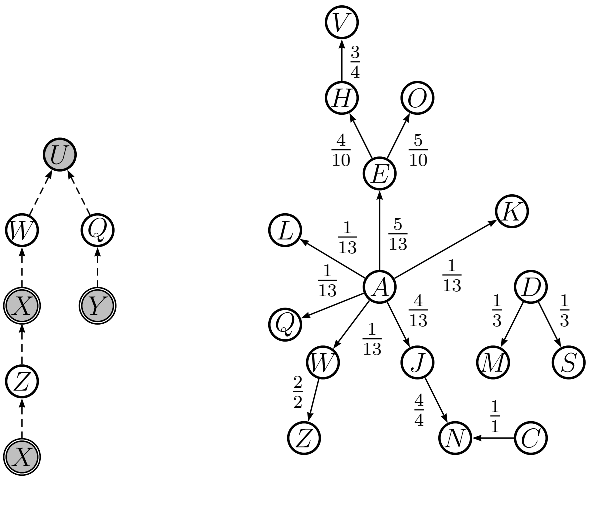
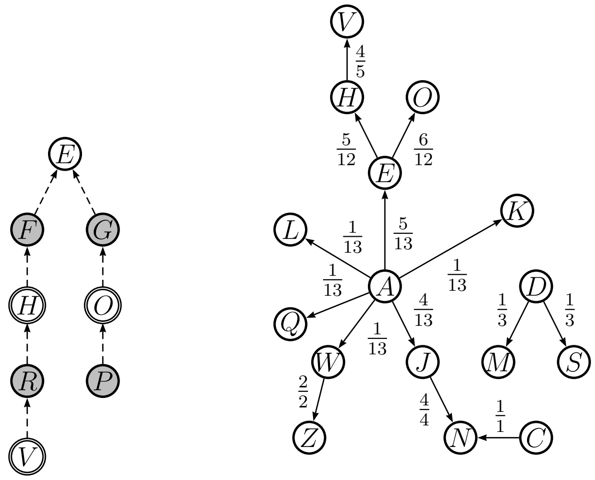


Figure 7.10: a) Dialogue $\mathscr{D}_9$, b) the correspondingly extended $\mathcal{R}G_1$



(a) (b)

Figure 7.11: a) Dialogue $\mathscr{D}_{10}$, b) the correspondingly extended $\mathcal{R}G_1$

$, < Ag_3, H >, < Ag_1, R >, < Ag_3, V >, < Ag_1, G >, < Ag_3, O >, < Ag_1, P >>$. *Winner of the dialogue is* Ag_1 *whose last move* (P) *results in* E *being labelled* **`out`**. *The dialogue along with the correspondingly extended* $\mathcal{R}G_1$ *is shown in Figure 7.11.*

***Step** 2:* **A dialogue between agents Ag_1 and Ag_6:**

Dialogue 11: *Ag_6 vs Ag_1, on the grounded acceptability of E. Arguments are moved in the following order:* $\mathscr{D}_{11} =<< Ag_6, A >, < Ag_1, B >, < Ag_6, E >, < Ag_1, F >, < Ag_6, H >, < Ag_1, D >, < Ag_6, J >, < Ag_1, U >, < Ag_6, W >, < Ag_1, Y >>$. *Winner of the dialogue is* Ag_1 *whose last move* (Y) *results in* A *being labelled* **`out`**. *The concerned dialogue appears in Figure 7.12.*

***Step** 3:* **Initial Opponent Model production:**

Based on the moves introduced into $\mathscr{D}_{11}$, Ag_1 *is able to construct an initial form of the OM of* Ag_6 *expressed as the set of arguments* $\mathcal{A}_{(1,6)} = \{A, E, H, J, W\}$ *as it appears in Figure 7.13a.*

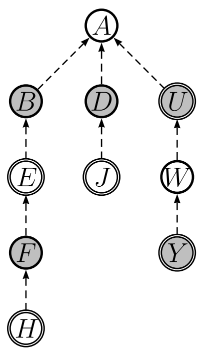


Figure 7.12: Dialogue 11, between agents Ag_6 (Pr) and Ag_1 (Op), on the grounded acceptability of A

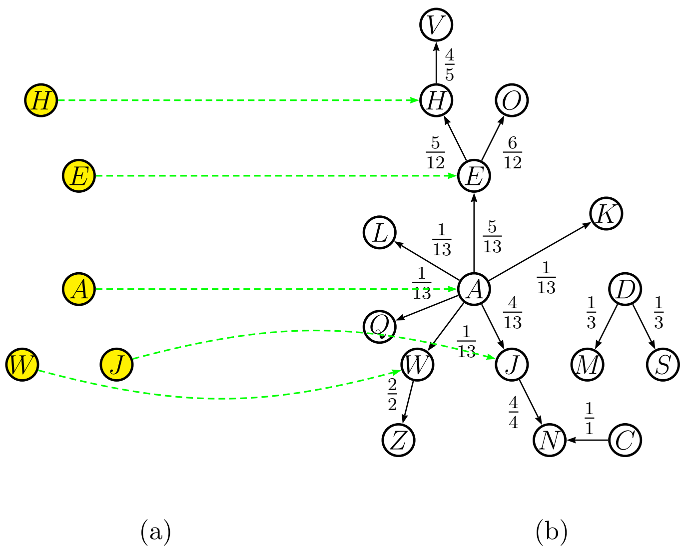
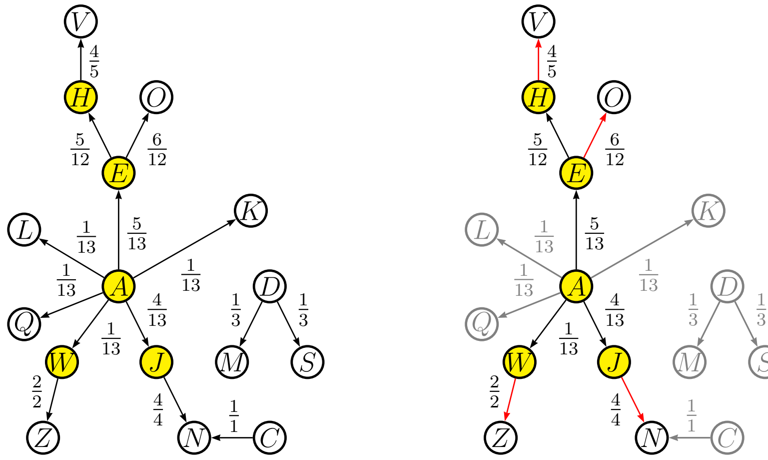


Figure 7.13: a) Set $\mathcal{A}_{(1,6)}$, b) $\mathfrak{R}G_1$



(a) (b)

Figure 7.14: a) The projection of $\mathcal{A}_{(1,6)}$ into $\mathcal{R}G_1$, b) The neighbouring nodes to the initial opponent model to be incorporated into $\mathcal{A}_{(1,6)}$

Step 4: The Augmentation Process:

Given the initial opponent model $\mathcal{A}_{(1,6)}$, and $\mathcal{R}G_1$, the modeller, Ag_1, can map the first into the second (i.e. project the initial OM, Figure 7.13a, into $\mathcal{R}G_1$, Figure 7.13b), and attempt to augment it with neighbours that satisfy a likelihood augmentation threshold greater or equal to 0.5.

As in this case the in-bound edges for all the neighbouring arguments $N_{\mathcal{A}} = \{V, O, K, N, Z, Q, L\}$ of set $\mathcal{A}_{(1,6)} = \{A, E, H, J, W\}$ is equal to 1, i.e. all neighbouring arguments of the yellow sub-graph are each of them connected to it with just one edge, it is obvious that the Monte-Carlo simulation would produce augmentation likelihoods for each of the elements of $N_{\mathcal{A}}$ with approximate numbers to the weight values on the arcs which link them to the elements of $\mathcal{A}_{(1,6)}$ (i.e. approximations of the weight values of the out-bound edges of the yellow sub-graph in Figure 7.14a).

Take for example the neighbouring argument N. N is connected to the OM (to the yellow nodes) with just one arc, r_{JN}. Therefore the prob-

ability/likelihood of including N, $Pr(N)$, is equal to w_{JN}. If however, argument C was also in the OM, i.e. if $C \in \mathcal{A}_{(1,6)}$, then $Pr(N)$ would be equal to the probability of including r_{JN} which is equal to w_{JN}, plus the probability of including r_{CN} which is equal to w_{CN}, minus their intersection, i.e:

$$Pr(N) = w_{JN} + w_{CN} - w_{JN} \cdot w_{CN}$$

which is obviously a lot more complex to solve (we refer the reader to Example 7 for a more detailed analysis). Since in this case no more than one edges connect each neighbouring argument to the OM, performing the simulation is redundant.

In this respect, it can be easily seen that the arguments which satisfy the augmentation threshold are $\{V, O, N, Z\}$ as illustrated in Figure 7.14b which will be used to augment $\mathcal{A}_{(1,6)}$, and produce the new set of arguments:

$$\mathcal{A}'_{(1,6)} = \mathcal{A}_{(1,6)} \cup \{V, O, N, Z\} = \{A, E, H, J, W, V, O, N, Z\}$$

assumed to be known to Ag_6.

Step** 5:* ***Accuracy comparison:

Provided Ag_6's actual knowledge $\mathcal{A}_{(6,6)}$ (Figure 7.15a), then, the `correctness` *of the augmented opponent model $\mathcal{A}'_{(1,6)}$ to the actual opponent knowledge $\mathcal{A}_{(6,6)}$ is:*

$$\varrho'_{(1,6)} = \frac{|\mathcal{A}'_{(1,6)} \cap \mathcal{A}_{(6,6)}|}{|\mathcal{A}'_{(1,6)}|} = \frac{8}{9}$$

its respective `completeness` *equal to:*

$$\omega'_{(1,6)} = \frac{|\mathcal{A}'_{(1,6)} \cap \mathcal{A}_{(6,6)}|}{|\mathcal{A}_{(6,6)}|} = \frac{8}{15}$$

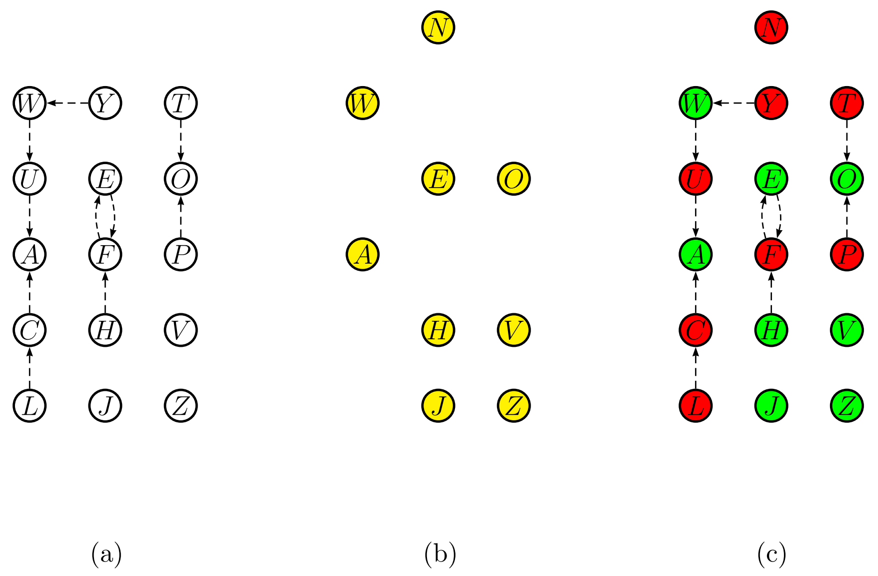


Figure 7.15: a) The set of arguments $\mathcal{A}_{(6,6)}$, b) The augmented OM $\mathcal{A}'_{(1,6)}$, c) Their accuracy: green nodes are matching arguments; red nodes are arguments with no match.

while $\mathcal{A}'_{(1,6)}$'s **`accuracy`** *to $\mathcal{A}_{(6,6)}$, is equal to:*

$$\hat{\alpha}'_{(1,6)} = \varrho'_{(1,6)} \cdot \omega'_{(1,6)} = \frac{8}{9} \cdot \frac{8}{15} = \frac{64}{135} \approx 0.47$$

In a similar sense, the **`correctness`** *of the initial opponent model $\mathcal{A}_{(1,6)} = \{A, E, H, W, J\}$, which is not augmented and relies only on the direct collection of information from the dialogues' commitment stores, to the actual opponent knowledge $\mathcal{A}_{(6,6)}$ is:*

$$\varrho_{(1,6)} = \frac{|\mathcal{A}_{(1,6)} \cap \mathcal{A}_{(6,6)}|}{|\mathcal{A}_{(1,6)}|} = \frac{5}{5}$$

its respective **`completeness`** *equal to:*

$$\omega_{(1,6)} = \frac{|\mathcal{A}_{(1,6)} \cap \mathcal{A}_{(6,6)}|}{|\mathcal{A}_{(6,6)}|} = \frac{5}{15}$$

while $\mathcal{A}_{(1,6)}$'s ***`accuracy`*** *to $\mathcal{A}_{(6,6)}$, is equal to:*

$$\hat{\alpha}_{(1,6)} = \varrho_{(1,6)} \cdot \omega_{(1,6)} = \frac{5}{9} \cdot \frac{5}{15} = \frac{25}{135} \approx 0.19$$

Finally, if this was a case study, a comparison between the two accuracy metrics $\hat{\alpha}'_{(1,6)}$ and $\hat{\alpha}_{(1,6)}$, would reveal that the augmented opponent model is more valid than the non-augmented one since:

$$\hat{\alpha}'_{(1,6)} > \hat{\alpha}_{(1,6)}$$

The above example illustrates a comparative evaluation, where an OM constructed based on conventional approaches, i.e. through information collection solely from an opponent's commitment store (in this case $\mathcal{A}_{(1,6)}$), is compared with its augmented version ($\mathcal{A}'_{(1,6)}$) in terms of which one is the most accurate, to the opponent's actual knowledge ($\mathcal{A}_{(6,6)}$). In relation to the typifications, which appear in Section 7.2, set $\mathcal{A}_{(1,6)}$ can be characterised as `Contained` as opposed to set $\mathcal{A}'_{(1,6)}$ which is `Partial`.

In this example we assumed a high level approach focussing on comparing the accuracy of the instantiated arguments of an opponent's sub-theory ($S_{(i,j)}$). However, one could focus on the sets of premises $\mathcal{K}$ and rules $\mathcal{R}$ of a sub-theory and compare those for a more thorough evaluation. For example, assuming a sub-theory:

$$S_{(1,6)} =< \mathcal{K}_{(1,6)}, \mathcal{R}_{(1,6)}, \leq'_{(1,6)}, \leq_{(1,6)}, \mathcal{G}_{(1,6)} >$$

and its augmented version:

$$S'_{(1,6)} =< \mathcal{K}'_{(1,6)}, \mathcal{R}'_{(1,6)}, \leq''_{(1,6)}, \leq'_{(1,6)}, \mathcal{G}'_{(1,6)} >$$

and given an opponent's actual knowledge:

$$S_{(6,6)} =< \mathcal{K}_{(6,6)}, \mathcal{R}_{(6,6)}, \leq'_{(6,6)}, \leq_{(6,6)}, \mathcal{G}_{(6,6)} >$$

one may produce and compare the accuracies of:

$$\mathcal{K}_{(1,6)} \ \ to \ \ \mathcal{K}_{(6,6)} \ \ with \ \ \mathcal{K}'_{(1,6)} \ \ to \ \ \mathcal{K}_{(6,6)}, \ \ and;$$

$$\mathcal{R}_{(1,6)} \ \ to \ \ \mathcal{R}_{(6,6)} \ \ with \ \ \mathcal{R}'_{(1,6)} \ \ to \ \ \mathcal{R}_{(6,6)}$$

Generally, provided an incrementally constructed $\mathcal{R}G$, built from a series of actual dialogue games, iterations of these comparisons may be performed based on dialogue interactions of the modeller with various opponents. In addition, one may perform a series of distinct tests to evaluate the added value of augmentation in cases where nothing is known about its opponent's (i.e. the initial opponent model is empty like in the case of our example) and in cases where an OM already contains some information.

Furthermore, relying on the results of such tests one may compare the effectiveness of producing valid OMs based on each of the different interrelatedness factors used for the construction of $\mathcal{R}G$s. In other words, to compare augmentations based on: support $\mathcal{R}G$s; common-attack-targets $\mathcal{R}G$s, and; based on both. In addition, the added value of producing support-$\mathcal{R}G$s with $\theta_t > 1$ could also be investigated. Finally, the evolution of the validity of a constantly augmented OM could also be investigated. In other words, researching whether a constantly augmented OM will eventually converge or deviate from the actual opponent knowledge.

7.4 Further Evaluations - A Discussion

Example 11 is concerned with an off-line evaluation approach, according to which the accuracy of an augmented OM to the opponent's actual knowledge is measured and compared with the conventional non-augmented OM, at the end of the dialogue process. However, one could attempt such an evaluation on the fly, i.e. during a dialogue process, through which, apart from measuring the accuracy of the incrementally constructed OM, the impact of augmenting an OM on the effectiveness of strategising could be put to the test.

For example, let us re-examine how dialogue $\mathscr{D}_{11}$ could evolve if the modeller Ag_1 opted to strategise while concurrently building and augmenting its model of

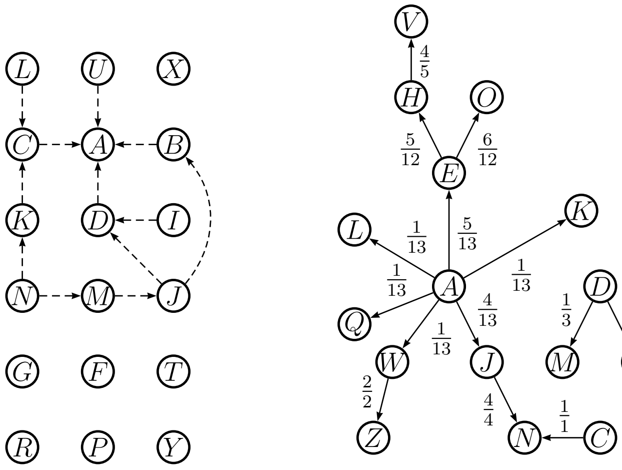
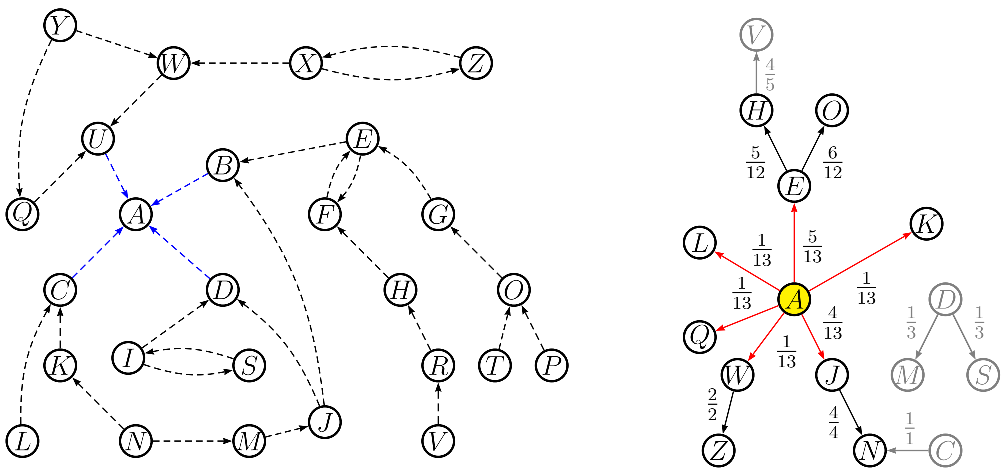


Figure 7.16: a) Ag_1's actual knowledge, b) Ag_1's relationships graph $\mathcal{R}G_1$

Ag_6, at the same time. Let us assume that Ag_1's actual knowledge represented as abstract arguments, $\mathcal{A}_{(1,1)}$, along with their binary attack relationships, is as it appears in Figure 7.16a. Note that the graph of Figure 7.16a is a sub-graph of the $O\mathcal{G}$ that appears in Figure 7.1. Let us also assume that Ag_1 is aware of the relationships graph $\mathcal{R}G_1$, constructed earlier through a series of 10 dialogues with other agents (Figure 7.16b).

Let these two agents engage in a dialogue for the grounded acceptability of argument A where Ag_6 is the proponent and Ag_1 the opponent. As soon as argument A is introduced into the game, Ag_1 is found at a strategic point where a choice has to be made between its possible options: arguments B, C, D and U (Figure 7.17a). Since no information is known so far about Ag_6's knowledge, Ag_1 can rely on the statistical information provided in $\mathcal{R}G_1$ to simulate the forest of the possible dialogue trees that could take place, and to instantiate a game tree from them, in order to decide on the next move.

Based on $\mathcal{R}G_1$ and given that A is obviously part of Ag_6's knowledge, Ag_1 anticipates that Ag_6 is likely to be able to support A with E and with a likelihood



(a) (b)

Figure 7.17: a) The omni-graph in which the possible attacks against A are highlighted in blue, b) The $\mathcal{R}G_1$ in which A's neighbouring edges appear in red.

of 5/13, with J and with a likelihood of 4/13, and with either L, Q, K or W with a likelihood of 1/13 (Figure 7.17b). This likelihoods can be passed as confidence values for their respective opponent arguments in the dialogue tree. Respectively, arguments E, J, L, Q, K and W are attacked themselves by moves known to the proponent (Ag_1): E is attacked by F and G, J by M, K by N, Q by Y, and W by X and Y. However, progressing to the next levels of the possible dialogue trees, solely based on the fact that A is the only argument certainty known to the opponent, requires that propagated likelihood values are computed for the arguments which could possibly follow after Ag_1's moves, accounting for their distance from A in $\mathcal{R}G_1$. For example, the probability, and respectively the confidence value, of getting to argument H from A depends on first getting to argument E and would therefore be equal to:

$$c_H = w_{AE} \cdot w_{EH} = \frac{5}{13} \cdot \frac{5}{12} = \frac{25}{156} \approx 0.16$$

Obviously these values will keep on diminishing as one moves further and further away from A in $\mathcal{R}G_1$. However, what we are stressing through this example is that even in the case where an agent knows nothing about its opponent he may rely on the $\mathcal{R}G$ in order to strategise against him. In other words, to simulate the forest of the possible dialogues based on which the actual dialogue could evolve, even through relying on propagated probability values. In this respect, one could evaluate the impact of augmentation on the effectiveness of strategising by performing numerous tests, where agents engage in dialogues against unknown opponents, differentiating between agents that use, and agents that do not use, $\mathcal{R}G$s, and comparing their dialogue outcomes.

7.5 Summary & Conclusions

The purpose of this chapter was the provision of a methodology towards evaluating the augmentation approach proposed in Chapter 4 used for the incorporation of additional information in an OM. We presented a summarising review of the current status-quo of argument-based benchmarking, which is only now being established through approaches such as `Probo`, `Tweety`, `TOAST` and DGEP, focussing on those that are able to account for structural argumentation frameworks and dialogues.

We provided two evaluation metrics concerned with the correctness and the completeness of an OM in an attempt to quantify the level of accuracy with which an OM may be characterised. The first concerns the level at which the information contained in a OM is also contained in the opponent's actual knowledge (correctness), while the second concerns the level at which the information contained in an opponent's actual knowledge is also contained in the OM (completeness). Of course, for the purpose of such an evaluation it is necessary that the an opponent's actual knowledge is known in advance.

In order to show how these metrics may be used for evaluation purposes we provided an extended evaluation example, through which a $\mathcal{R}G$ was instantiated by a certain agent against multiple other agents, and which was later used against a certain opponent to augment knowledge acquired directly from a single dialogue against him. At the end of the example, the accuracy of the augmented OM was compared with that of the non-augmented one, in order to test the impact of the augmentation in increasing its accuracy.

We finally presented a discussion where we elaborated on how a $\mathcal{R}G$ can be used on the fly for dialogue simulation and strategising purposes, suggesting another approach to evaluating augmentation; one that is concerned with its impact in producing a more effective strategising.

The evaluation of our approach is certainly an important part of our research, as it is for every research. And this is not just so that our research approach can be deemed successful or not, through appropriate testing, but also because evaluations often reveal aspects of modelling, or the impact of accounted factors through revealing their consequences, in ways and that were often not anticipated. This information can lead to the optimisation of our approach as well as to the investigation of other issues related with opponent modelling, such as accounting for domain or context related information, the perception of general knowledge and group-only accessed knowledge, in combination with the properties of the relationship graphs produced in each case etc., and it is thus our immediate future research objective.

Furthermore, as, apart from the interrelatedness of information, linking of arguments could to some extent rely on how groups of agents in a multi-agent environment share common information, leading to certain arguments following more often against a certain group of agents but not against others, an interesting direction would be to extend the dialogue framework proposed in this thesis in order incorporate additional information accounted by social argumentation frameworkss (SAFs) (Leite and Martins [2011]) that we currently ignore. We anticipate that the added information will very likely lead to an even better quantification of the likelihood relationships between opponent arguments in a $\mathcal{R}G$ and we thus intend to also put this assumption to the test in our future work.

Chapter 8

Conclusions & Future Work

In this final chapter we summarise our work listing the research gaps that we have addressed as well as our corresponding contributions. In addition we list the limitations in our work and offer a number of open problems for investigation towards which we intend to turn in our future work.

8.1 Summary of Work & Contributions

In this thesis we addressed a number of research questions defined in Section 1.2. For the purpose of exploring the impact of the underlying logic in dialogues and the development of an dialogue framework for structured arguments (questions 1 & 2) we developed an $ASPIC^+$-based framework for persuasion dialogue. $ASPIC^+$ has been shown to satisfy Caminada's [2007] rationality postulates, it explicitly models the logical content and structure of arguments, while it accommodates many existing logical approaches to argumentation. The latter allows us to claim a similar level of generality for our dialogical framework.

Relying on this framework, we showed that it is possible that new arguments are instantiated from the logical information introduced through the exchanged arguments in a dialogue. This fact impacts both strategy development in dialogues as well as argument evaluation protocols in dialogue frameworks that guarantee soundness and completeness.

Particularly in relation to strategic considerations, we further illustrated how in combination with the existence of an OM, the possible instantiation of arguments, *not just from within the exchanged information*, but also partly from the opponent's possible knowledge is also something that should be taken into account. If one is rational, revealing information to an opponent that would allow her to construct arguments which could reinstate the opponent's position is something that should be avoided.

In relation to the 3^{rd} question:

- How will the employment of a structured argumentation system add to the expressiveness of dialogues produced in this framework, allowing for the introduction of a participant's preferences as a means of justifying their rational with respect to a defeat relationship?

Within the context of the $ASPIC^+$-based framework, we were able to allow agents to introduce preferences against arguments under two forms of attacks: preference attack, which concerns the attack of a preference on an argument, and; preference-rebut attacks which concerns the attack of preference against another preference. This way we allowed participants to maintain their individual preferences in a game, as well as to use them in order to argue about their rational on the defeat relationship between contradictory arguments.

In relation to the 4^{th} and the 5^{th} question:

- How can we evaluate the dialogue outcomes produced by our framework with respect to the acceptability of a disputed argument?

- With what restrictions should protocols developed for our dialogue framework be characterised with, so as to guarantee the soundness of a dialogue's result with respect to those semantics?

We developed two protocols, drawing from the work of Modgil and Caminada [2009] for the credulous preferred semantics as well as for the grounded semantics. For dialogues characterised by these protocols we were able to provided soundness and fairness results, in the case where the dialogues were satisfying conditions concerned with logical (Prakken and Vreeswijk [2002]) and protocol completeness. In simple words, the first requires that at a given point in a dialogue all moves that could be instantiated from the accumulated knowledge implicitly constructed from the exchange of arguments have been moved into the game. The second requires that all arguments that could have been repeated in the game have been repeated. These results prove a correspondence between the production of a justified argument in dialogues that satisfy certain properties in our system, with the participation of an argument in an extension of the AF implicitly constructed from those dialogues.

We also offered a preference conflict resolution method, concerned with the set of the accumulated preferences introduced into a game, in a way that reflected the employed semantics in each game.

In relation to the 6^{th} question:

- What are the main mechanisms based on which one may build, update and maintain an OM?

We provided a modelling framework for building, updating and maintaining an OM, which assumes a structured representation of knowledge in the form of logical constituents for argument instantiation. The framework allows the modeller to differentiate between 3 ways of opponent information acquisition, those of: direct collection of information through an opponent's commitment store; collection from third parties, and; augmentation.

In relation to the 7^{th} and the 8^{th} question:

- What factors affect the credibility of acquired opponent information?

- Can we formalise an opponent modelling process based on which one could account for these factors in a way that could possibly increase the effectiveness of a strategy?

We identified factors related to trust issues in the case of third party provided information, while in the case of augmentation these factors concern the frequency in which an argument appears in dialogues with other arguments and with respect to some interrelatedness between the presumably related arguments. We were able to quantify the impact of these factors in the form of confidence values and associate acquired information with such values based on their information acquisition method. We also showed how these values can be used for utility evaluation purposes, assuming the application of the minimax algorithm.

In relation to the 9^{th} question:

- Can we develop a modelling mechanism for inducing and augmenting an OM based on a modeller's general experience in dialogues?

We proposed an augmentation process, and developed a corresponding mechanism which allows a modeller to incorporate additional information (external to a current OM) from its general experience, into an OM, after relating it with information already in the model, thus utilising its experience in a multifaceted way. Specifically, the mechanism relies on the interrelatedness of information, captured by notions of support and that of common attack targets between the arguments, for instantiating a relationships graph ($\mathcal{R}G$). We showed that one can then augment an OM by mapping the arguments instantiated from it in the $\mathcal{R}G$ to include neighbouring arguments (their logical constituents) into the model contingent on satisfaction of some likelihood threshold.

In relation to the 10^{th} question:

- If the proposed modelling process is not tractable, resulting in making strategy development processes also intractable, how can we increase its tractability?

The process we proposed was shown to have a high complexity with respect to the computation of the likelihoods of the possible augmentations of an OM. In order to allow for it to be used in dialogues on the fly for strategising purposes, we developed a Monte-Carlo simulation for the approximate computation of those likelihoods. We evaluated our approach using randomly generated graphs, showing the rate of convergence of our sampling to the theoretical values, while we proved that our approach converges after a constant number of steps.

Finally, in relation to the 11^{th} question:

- How can we evaluate the effectiveness of an opponent modelling mechanism in increasing a model's validity and consequently its credibility?

We defined two metrics respectively related to the correctness and the completeness of an OM, which are essentially concerned with the correspondence between an opponent's OM and the opponent's actual knowledge. Given the lack of argument-based benchmarks for dialogues we provided a methodology towards evaluating our approach, and left the actual evaluation to future work.

8.2 Limitations & Future work

Through our work, we were able to identify a number of limitations. We list these limitations in this section as open problem for future investigation.

1. **Update of an Agent's own Knowledge:**

   An issue left unexplored in this thesis is how agents may update their own knowledge $S_{(i,i)}$ after the acquisition of new information from the dialogue process. For the purpose of this thesis we simply assume that regardless of whether new information is introduced in an agent's theory, no information is retracted but rather conflicts are resolved through evaluation of the justified arguments under acceptability semantics.

   However, apart from how agents update their own knowledge in dialogues, one has to also wonder about whether an agent's own knowledge should also contain the knowledge assumed to be known to all its opponents. As Rienstra et al. [2013] explain it is not reasonable for a modeller to believe that a certain opponent is aware of an argument without the modeller being aware of the argument himself. They refer to this as *"awareness restriction"*.

   On the same topic, we also do not account for how agents may update their own preferences in dialogues nor we account for how priority-orderings may be formed. We simply assume that agents use generic principles to do so such as the specificity principle, and the temporal principle which orders newly acquired knowledge over older knowledge.

2. **Modelling Goals:**

   Another issue left implicit in our work is the modelling of goals. Goals are first introduced in this thesis in the general dialogue framework of Chapter 3, where a general strategy function is also introduced. Although the concept of goals is not used here, any concrete implementation of the theoretical concepts discussed in this thesis will require reference to an agent's goals, and thus accounting for goals in our framework was imperative, at least for the sake of theoretical completeness.

There are many objectives that an agent may be pursuing at any given time in a dialogue. Agents may decide to commit to their dialogical objectives corresponding to their roles in dialogues, or aim towards maximising their personal utility, which implies that they may have to deviate from those roles. Apart from listing a possible set of objectives that may characterise an agent's goals, we also intend to research how one may develop ways of modelling strategies that correspond to those goals, e.g. as an automaton (Carmel and Markovitch [1998]), as well as mechanisms used to anticipate the use of such strategies by an opponent in a dialogue. One can then rely on these mechanism for counter-strategising purposes.

3. **Additional Factors that may Affect the Quantification of likely Related Arguments in a $\mathcal{RG}$:**

   Though in our work we essentially account for the frequency of appearance of pairs of arguments related in different ways in a history of dialogues for deducing the relationship likelihood between them, other factors could be taken into account as well. Such factors may concern issues related with the context of the game, or the assumed background of the participants, or even whether the participants are members of communities with access to shared knowledge.

   Specifically in relation to the latter, an idea could be to instantiate a network of agents, with agents being linked and with those links being strengthened with weights, every time those agents use the same arguments. One may then apply a community detection algorithm for deducing communities of agents, in a multi-agent environment which could share the same arguments.

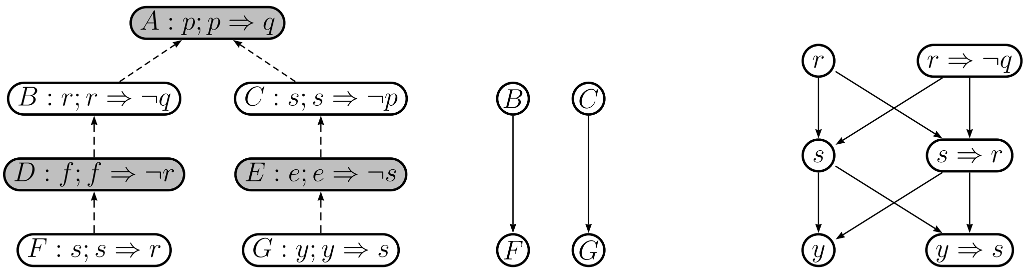
   Furthermore, an interesting issue worth exploring is the extent to which the proposed quantification methods are susceptible to manipulation. For instance, in the case of thw weighting mechanism proposed in Definition 52, in order to diminish the significance of two related arguments $A \rightarrow B$ in a $\mathcal{RG}$ a player may invoke A many times, leaving a player anxious to maintain the weighting w_{AB} only with the option of responding B every time A appears. Of course, this stance would require that one commits to only responding with B sacrificing the use of alternatives which may possibly lead to outcomes with higher utility. This and other such issues will be part of our future work.

4. **Modelling of a logical constituents $\mathcal{R}G$:**

   Another future research objective is concerned with attempting a more analytic approach of an instantiated $\mathcal{R}G$, where the logical constituents of arguments could be linked instead of the arguments they compose.

   This objective derives from a limitation concerned with the possible ways of calculating the confidence value of a constructed argument, discussed in Section 4.5.2 of Chapter 4. To recap, we proposed three possible ways of doing so: through propagating the confidence values of the argument's constituents; computing their average, or; choosing amongst them the one with the smallest value to pass to the constructed argument. As explained, in the case of propagation we were faced with an asymmetry issue. Namely, if confidence values are assigned to the constituents of arguments prior to their incorporation into an OM, then the reconstructed confidence value of that argument after it has been re-instantiated from an OM, should match its original value. However this is not the case if propagation is used. This asymmetry does not appeal to intuition.

   This problem results from treating the arguments, and particularly the possibility of their awareness by a certain opponent, as random events instead of doing so for their constituents. After all, instantiation of an argument is dependent on whether the logical constituents that compose that argument are known to a certain opponent. So, when a modeller enquires as to whether an opponent knows a certain argument, the actual question should be "is that opponent capable of constructing that argument?" or "what is the likelihood of that opponent to be aware of certain constituents which make the construction of a certain argument possible?" (this approach relates to work by Riveret et al. [2007]).



(a) (b) (c)

Figure 8.1: a) A dialogue tree $\mathcal{T}$, b) The resulting $\mathcal{R}G$, c) A relationships graph of the related constituents

In other words, from the scope of probability theory, the random events should be the likelihood of an opponent to be aware of the constituents of certain arguments, while arguments should simply be characterised by the result of their propagation.

For doing so it is required that we provide a way for modelling relationships between constituents in a logical $\mathcal{R}G$, rather than using an abstract one which relies on relationships between arguments. The notion of support can still serve as a connection component. However, instead of connecting arguments we will be connecting their constituents. This idea is illustrated in Figure 8.1.

If we adhere to the support relationship conditions imposed by Definition 51 then the resulting $\mathcal{R}G$ for the opponent arguments used, would be the one that appears in Figure 8.1b based on which one could deduce a structural instantiation of it which appears in Figure 8.1c. The added expressiveness of this graph could possibly better reflect relationships between arguments not captured by an 'abstract' $\mathcal{R}G$.

5. **Development of benchmarks for dialogues, to test the proposed modelling implementation:**

   The augmentation mechanism proposed in this thesis, relies on features concerned with the interrelatedness of information and draws from the structure of dialogues to statistically infer possible opponent knowledge. Though it intuitively appears to be a reasonable way of anticipating opponent information and exploiting a modeller's general dialogue experience, it remains untested.

As discussed in Chapter 7 this is because of the lack dialogue-based argumentation benchmarks. Even now, at the time this thesis is written, most argument-based benchmarks are concerned argument evaluation (Snaith and Reed. [2012]; Thimm [2014]), and to the best of our knowledge there is very little work on the development of argument-based dialogue benchmarks (Bex et al. [2014]).

Another problem is that anticipating possible opponent knowledge based on statistical inference in a sense reflects how information is distributed amongst the members of a studied group. In other words, for the purpose of practically evaluating our approach, actual knowledge needs to be collected and represented in the form of argument graphs, in order to guarantee unbiased results. This process is referred to as *argument mining* and it is only now gaining research interest as a new challenge in corpus-based discourse analysis.

However, simulating such knowledge in the form of argument graphs for evaluation purposes is still feasible. That is, as long as it satisfies certain structural properties that argument graphs are expected to have (Hunter and Woltran [2013]). The evaluation methodology presented in Chapter 7 relies on such graphs, and the methodology proposed relies on a scenario according to which an augmented opponent model is compared against the opponent's actual knowledge, to test whether a better match has been created. Our immediate research direction is to develop similar dialogue scenarios, which are theoretically mapped out in terms of agents' knowledge evolving and how the augmentation would work, which will become a benchmark to test our augmentation approach.

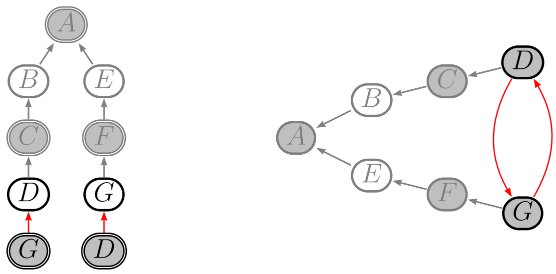
Appendix A

Miscellaneous

A.1 Proofs

Proposition 2 Let $\mathcal{T}$ be a dialogue tree and $\mathcal{DM}_0$ be $\mathcal{T}$'s root move, if $\mathcal{DM}_0$ is labelled `in` then it is not necessary that $\exists \mathcal{T}' \in \mathcal{T}$.

Proof Figure A.1 presents a counter example which serves as proof, for a credulous game between two agents.



(a) (b)

Figure A.1: a) A dialogue tree $\mathcal{T}$ (Grey moves by Pr), b) The AF induced from $\mathcal{T}$.

The example is based on the AF presented in [Modgil and Caminada, 2009, p. 125] which deals with the exact same problem but in the concept of arguments games. In an analogous way, the concerned AF (A.1b) can be induced from the terminated credulous dialogue presented in Figure A.1a. Notice that:

1. The root argument A is labelled `in`, since the end notes in both disputes are Pr arguments.
2. As Op cannot repeat its moves in the credulous game, Pr is the winner of this game.
3. However, the induced AF presented in A.1b reveals a rebut attack relationship between arguments G and D, both of which are moved by Pr.
4. 3 implies that the set of arguments moved by Pr is not conflict free.

A.2 Nested Opponent Models

Definition 71 (Agent Theory with nested OMs) *Let $\{Ag_1, \ldots, Ag_\nu\}$ be a set of agents. For $i = 1 \ldots \nu$, the theory of Ag_i is a tuple $AgT_i =< S_{(i,1)}, \ldots, S_{(i,\nu)} >$ such that:*

For $j = 1 \ldots \nu$, each sub-theory $S_{(i,j)} = \langle AT_{(i,j)}, \mathcal{G}_{(i,j)}, S'_{(j,i)} \rangle$ where:

- *$AT_{(i,j)}$ is what Ag_i believes is the argumentation theory $(AS_{(i,j)}, KB_{(i,j)}, p_{(i,j)})$ of Ag_j;*
- *$\mathcal{G}_{(i,j)}$ is what Ag_i believes are the goals of Ag_j;*
- *$S'_{(j,i)} = \langle AT'_{(j,i)}, \mathcal{G}'_{(j,i)}, S'_{(i,j)} \rangle$ is what Ag_i believes is Ag_j's OM of Ag_i;*
- *If $j = i$, then $AT_{(i,j)}$ and $\mathcal{G}_{(i,j)}$ are respectively Ag_i's own argumentation theory and goals, while $S'_{(i,j)} = \varnothing$ (null tuple);*

Then for $i, j, k, m = 1 \ldots n$, let:

$$S_{(i,j)} = \langle AT_{(i,j)}, \mathcal{G}_{(i,j)}, S'_{(j,i)} \rangle, \quad and \quad S_{(k,m)} = \langle AT_{(k,m)}, \mathcal{G}_{(k,m)}, S'_{(m,k)} \rangle$$

be any two distinct sub-theories, where:

$$AT_{(i,j)} = (AS_{(i,j)}, KB_{(i,j)}, p_{(i,j)}), \ and \quad AT_{(k,m)} = (AS_{(k,m)}, KB_{(k,m)}, p_{(k,m)})$$

		1	2	3	4	5	6
		$\mathcal{K}$	$\leq'$	$\mathcal{R}$	$\leq$	$\mathcal{G}$	$S'_{(j,i)}$
1	$S_{(i,1)}$	$\mathcal{K}_{(i,1)}$	$\leq'_{(i,1)}$	$\mathcal{R}_{(i,1)}$	$\leq_{(i,1)}$	$\mathcal{G}_{(i,1)}$	S'_{1_i}
2	$S_{(i,2)}$	$\mathcal{K}_{(i,2)}$	$\leq'_{(i,2)}$	$\mathcal{R}_{(i,2)}$	$\leq_{(i,2)}$	$\mathcal{G}_{(i,2)}$	$S'_{(2,i)}$
.	...	...	...	...	...	...	...
i	$S_{(i,i)}$	$\mathcal{K}_{(i,i)}$	$\leq'_{(i,i)}$	$\mathcal{R}_{(i,i)}$	$\leq_{(i,i)}$	$\mathcal{G}_{(i,i)}$	$\varnothing$
.	...	...	...	...	...	...	...
ν	$S_{(i,\nu)}$	$\mathcal{K}_{(i,\nu)}$	$\leq'_{(i,\nu)}$	$\mathcal{R}_{(i,\nu)}$	$\leq_{(i,\nu)}$	$\mathcal{G}_{(i,\nu)}$	$S'_{(\nu,i)}$

Table A.1: The distinct sets of logical elements found in each sub-theory of a single agent theory (AgT_i) extended to include nested OMs

as explained in Section 3.1.1, we then assume that:

- $p_{(i,j)} = p_{(k,m)}$*, where* $p_{(i,j)}$ *and* $p_{(k,m)}$ *are the preference functions (used for defining an ordering over the set of all arguments that can be constructed from a KB in an AS) in* $AT_{(i,j)}$ *and* $AT_{(k,m)}$ *respectively;*
- $\mathcal{L}_{(i,j)} = \mathcal{L}_{(k,m)}$*, and* $\overline{\ }_{(i,j)} = \overline{\ }_{(k,m)}$*, where* $\mathcal{L}_{(i,j)}$ *(*$\overline{\ }_{(i,j)}$*) and* $\mathcal{L}_{(k,m)}$ *(*$\overline{\ }_{(k,m)}$*) are the languages (contrary relations) in* $AS_{(i,j)}$ *and* $AS_{(k,m)}$ *respectively.*

We should note that in the case where the modelling framework presented in chapter 4, was extended so as to account for nested OMs it is easy to see that an agent Ag_i may simply assume the role of its opponent Ag_j, and that relying on the same mechanism, described in Definition 44, Ag_j will also update its model of Ag_i based on Ag_i's commitment store (CSi) and so on. In any case, this *does not* mean that Ag_i's nested OM will accurately reflect its opponent's OM, since Ag_i cannot be aware of additional information that may be provided, for example, to Ag_j by an external source (a third party agent).

Table A.1 illustrates the sets of logical elements found in each sub-theory of a single agent theory, with the additional inclusion of nested OMs expressed as $S'_{(j,i)}$. Essentially $S'_{(j,i)}$ is what Ag_i believes Ag_j believes about Ag_i. Obviously this recursive modelling allows for numerous levels of opponent modelling. We differentiate between different levels of recursive modelling in a similar sense to how this is done by Carmel and Markovitch [1996], i.e.:

- a *0 level* model concerns a sub-theory where $S_{(i,j)} = \langle AT_{(i,j)}, \mathcal{G}_{(i,j)}, \varnothing \rangle$;
- a *1 level* model concerns a sub-theory where $S_{(i,j)} = \langle AT_{(i,j)}, \mathcal{G}_{(i,j)}, S'_{(j,i)} \rangle$ and $S'_{(j,i)}$ is a 0 level recursive model $S'_{(j,i)} = \langle AT'_{(j,i)}, \mathcal{G}'_{(j,i)}, \varnothing \rangle$, and;
- a *n-level* model concerns a sub-theory where $S_{(i,j)} = \langle AT_{(i,j)}, \mathcal{G}_{(i,j)}, S'_{(j,i)} \rangle$, where $S'_{(j,i)}$ is a n-1 level recursive model.

A.3 Trust Network & Propagated Trust Computation

Definition 72 (Trust Network) *Let $Ags = \{Ag_1, \ldots, Ag_\nu\}$ be a set of agents in a multi-agent environment, then a trust network is a pair:*

$$T_N = \{Ags, \tau\}$$

where $\{\tau\}$ is the set of pairwise trust relations over the agents in Ags so that if $(Ag_i, Ag_j) \in \tau$ then $\{Ag_i, Ag_j\}$ is a directed arc in T_N.

Definition 73 (Computation of trust) *Let $p(Ag_i, Ag_j) =< Ag_i, Ag_{i+1}, \ldots, Ag_{j-1}, Ag_j >$ be a path between agents Ag_i and Ag_j such that:*

$$p(Ag_i, Ag_j) = \tau(Ag_i, Ag_{i+1}), \tau(Ag_{i+1}, Ag_{i+2}), \ldots, \tau(Ag_{j-1}, Ag_j)$$

and where the trust values between the agents is provided by a function tr such that:

- $tr^{[0,1]} : Ags \times Ags \mapsto [0, 1]$
- $tr(Ag_i, Ag_i) = 1$
- $tr(Ag_i, Ag_j) \neq 0 \Leftrightarrow (Ag_i, Ag_j) \in \tau$
- $tr(Ag_i, Ag_j) = 0 \Leftrightarrow (Ag_i, Ag_j) \notin \tau$

then, Ag_i trust in Ag_j is computed as follows:

$$tr(Ag_i, Ag_j) = tr(Ag_i, Ag_{i+1}) \otimes^{tr} tr(Ag_{i+1}, Ag_{i+2}) \otimes^{tr} \cdots \otimes^{tr} tr(Ag_{j-1}, Ag_j)$$

for some function $\otimes^{tr}$.

Bibliography

L. Amgoud and C. Cayrol. Integrating preference orderings into argument-based reasoning. In D. Gabbay, R. Kruse, A. Nonnengart, and H. Ohlbach, editors, *Qualitative and Quantitative Practical Reasoning*, volume 1244 of *Lecture Notes in Computer Science*, pages 159–170. Springer Berlin Heidelberg, 1997. 24, 25

L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3): 197–215, 2002a. 51, 94

L. Amgoud and C. Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *Journal of Automated Reasoning*, 29(2):125–169, 2002b. 140

L. Amgoud and N. Maudet. Strategical considerations for argumentative agents (preliminary report). In *Proceedings of the 9th International Workshop on Non-Monotonic Reasoning*, pages 399–407, 2002. 21, 22, 27

L. Amgoud and S. Parsons. Agent dialogues with conflicting preferences. In *Proceedings of the Eigth International Workshop on Intelligent Agents VIII*, pages 190–205, 2001. 25

L. Amgoud, L. Bodenstaff, M. Caminada, P. McBurney, S. Parsons, H. Prakken, J. Van Veenen, and G. Vreeswijk. Final review and report on formal argumentation system. Technical report, Deliverable D2.6, ASPIC IST-FP6-002307, 2006. 49, 65, 69, 70

J. L. Austin. *How To Do Things with Words*. Oxford University Press, 1962. 69

P. Baroni and M. Giacomin. Semantics of abstract argument systems. In G. Simari and I. Rahwan, editors, *Argumentation in Artificial Intelligence*, pages 25–44. Springer US, 2009. 39, 85

T. J. Bench-Capon, S. Doutre, and P. E. Dunne. Audiences in argumentation frameworks. *Artificial Intelligence*, 171(1):42 – 71, 2007. 23, 25, 79, 140

T. J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Logic and Computation*, 13(3):429–448, 2003. 141

T. J. M. Bench-Capon and P. E. Dunne. Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10-15):619–641, 2007. 103

T. J. M. Bench-Capon, S. Doutre, and P. E. Dunne. Asking the right question: forcing commitment in examination dialogues. In *Proceedings of the Second International Conference on Computational Models of Argument (COMMA'08)*, pages 49–60, 2008. 58

P. Besnard and A. Hunter. *Elements of Argumentation*. The MIT Press, 2008. 51

F. Bex, J. Lawrence, and C. Reed. Generalising argument dialogue with the dialogue game execution platform. In *Proceedings Fifth International Conference on Computational Models of Argument (COMMA'14)*, pages 141–152, 2014. 259, 260, 295

S. Bistarelli, F. Rossi, and F. Santini. A first comparison of abstract argumentation systems: A computational perspective. In *Proceedings of the Fifth International Conference on Corpus Linguistics (CILC'13)*, pages 241–245, 2013. 259

E. Black and K. Atkinson. Choosing persuasive arguments for action. In *Proceedings of the Tenth International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'11)*, pages 905–912, 2011. 21, 22, 26, 29

E. Black and A. Hunter. An inquiry dialogue system. *Autonomous Agents and Multi-Agent Systems*, 19:173–209, 2009. 21, 23

A. Bondarenko, P. M. Dung, R. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93(1–2):63–101, 1997. 49, 50, 54

G. Brewka. *Non-monotonic Reasoning: Logical Foundations of Commonsense*. The Press Syndicate of the University of Cambridge, 1991. 18, 46

M. Caminada. On the issue of reinstatement in argumentation. In *Proceedings of the Tenth European Conference on Logics in Artificial Intelligence (JELIA'06)*, pages 111–123, 2006. 19, 40

M. Caminada. An Algorithm for Computing Semi-Stable Semantics. In *Proceedings of the Ninth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)*, pages 222–234, 2007. 31, 40, 49, 54, 62, 287

M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007. 49

D. Carmel and S. Markovitch. Incorporating opponent models into adversary search. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence and Eighth Innovative Applications of Artificial Intelligence Conference, (AAAI'96, IAAI'96)*, volume 1, pages 120–125, 1996. 21, 22, 26, 28, 29, 146, 236, 237, 238, 298

D. Carmel and S. Markovitch. Model-based learning of interaction strategies in multi-agent systems. *Journal of Experimental and Theoretical Artificial Intelligence*, 10(3):309–332, 1998. 21, 22, 26, 29, 292

C. Cayrol, S. Doutre, and J. Mengin. On decision problems related to the preferred semantics for argumentation frameworks. *Logic and Computation*, 13 (3):377–403, 2003. 86

F. Cerutti, N. Oren, H. Strass, M. Thimm, and M. Vallati. A benchmark framework for a computational argumentation competition. In *Proceedings of the Fifth International Conference on Computational Models of Argument (COMMA'14)*, pages 459–460, 2014. 256, 259

C. Chesñevar, J. McGinnis, S. Modgil, I. Rahwan, C. Reed, G. Simari, M. South, G. Vreeswijk, and S. Willmott. Towards an argument interchange format. *Knowledge Engineering Review*, 21(4):293–316, 2006. 260

H. de Weerd, R. Verbrugge, and B. Verheij. How much does it help to know what she knows you know? An agent-based simulation study. *Artificial Intelligence*, 199-200:67–92, 2013. 237

P. M. Dung. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995. 18, 23, 37, 38, 39, 40, 46, 48, 53, 62, 113, 142

P. M. Dung, R. A. Kowalski, and F. Toni. Assumption-based argumentation. In *Argumentation in Artificial Intelligence*, pages 199–218. Springer, 2009. 258

P. Dunne and P. McBurney. Concepts of optimal utterance in dialogue: Selection and complexity. In F. Dignum, editor, *Advances in Agent Communication*, volume 2922 of *Lecture Notes in Computer Science*, pages 310–328. Springer Berlin Heidelberg, 2004. 21, 143

P. E. Dunne. Prevarication in dispute protocols. In *Proceedings of the Ninth International Conference on Artificial Intelligence and Law (ICAIL'03)*, pages 12–21. ACM Press, 2003. 57

P. E. Dunne. Suspicion of hidden agenda in persuasive argument. In *Proceedings of the First International Conference on Computational Models of Argument (COMMA'06)*, pages 329–340, 2006. 57

P. E. Dunne and T. J. M. Bench-Capon. Two party immediate response disputes: Properties and efficiency. *Artificial Intelligence*, 149(2):221–250, 2003. 43, 93, 94

P. E. Dunne and T. J. M. Bench-Capon. Identifying audience preferences in legal and social domains. In *Proceedings of the Fifteenth International Conference on Database and Expert Systems Applications (DEXA'04)*, pages 518–527, 2004. 140

P. E. Dunne and P. McBurney. Optimal utterances in dialogue protocols. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'03)*, pages 608–615, 2003. 20

P. E. Dunne, S. Doutre, and T. J. M. Bench-Capon. Discovering inconsistency through examination dialogues. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI'05)*, pages 1680–1681, 2005. 56

R. Eijk. *Programming Languages for Agent Communications*. PhD thesis, Department of Computer Science, Utrecht University, Utrecht, The Netherlands, 2000. 59

C. D. Emele, T. J. Norman, and S. Parsons. Argumentation strategies for plan resourcing. In *Autonomous Agents and Multi-Agent Systems*, pages 913–920, 2011. 21, 22, 30

X. Fan and F. Toni. Assumption-based argumentation dialogues. In *Proceedings of the Twenty-second International Joint Conference on Artificial Intelligence (IJCAI'11)*, pages 198–203, 2011. 21, 23

X. Fan, R. Craven, R. Singer, F. Toni, and M. Williams. Assumption-based argumentation for decision-making with preferences: A medical case study. In *Proceedings of the Fourteenth International Workshop on Computational Logic in Multi-Agent Systems (CLIMA XIV)*, pages 374–390, 2013. 32

D. M. Gabbay and J. Woods. More on non-cooperation in dialogue logic. *Logic Journal of the IGPL*, 9(2):305–324, 2001. 56

P. J. Gmytrasiewicz, E. H. Durfee, and D. K. Wehe. A decision-theoretic approach to coordinating multiagent interactions. In *International Joint Conference on Artificial Intelligence*, volume 1 of *IJCAI'91*, pages 62–68. Morgan Kaufmann Publishers Inc., 1991. 237

T. Gordon, H. Prakken, and D. Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171:875–896, 2007. 49

C. Hadjinikolis, S. Modgil, E. Black, and P. McBurney. Mechanisms for Opponent Modelling. In A. V. Jones, editor, *2012 Imperial College Computing Student Workshop*, volume 28 of *OpenAccess Series in Informatics (OASIcs)*, pages 62–68. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2012a. 34

C. Hadjinikolis, S. Modgil, E. Black, P. McBurney, and M. Luck. Investigating strategic considerations in persuasion dialogue games. In *Proceeding of the Starting AI Researchers' Symposium (STAIRS'12)*, pages 137–148. IOS Press, 2012b. 35

C. Hadjinikolis, Y. Siantons, S. Modgil, E. Black, and P. McBurney. Opponent modelling in persuasion dialogues. In *Proceedings of the Twenty-third International Joint Conference on Artificial Intelligence*, IJCAI'13, pages 164–170. AAAI Press, 2013. 35

C. L. Hamblin. *Fallacies*. Methuen, 1970. 74

M. Hazewinkel, editor. *"Minimax principle", Encyclopedia of Mathematics*, chapter 2. Springer, 2001. 253

A. Hunter. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, 54(1):47–81, 2013. 207

A. Hunter and S. Woltran. Structural properties for deductive argument systems. In L. van der Gaag, editor, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 7958 of *Lecture Notes in Computer Science*, pages 278–289. Springer Berlin Heidelberg, 2013. 258, 295

H. Jakobovits and D. Vermeir. Dialectic semantics for argumentation frameworks. In *Proceedings of the Seventh International Conference on Artificial Intelligence and Law (ICAIL'99)*, pages 53–62. ACM Press, 1999. 72

Y. Katz and J. Golbeck. Social network-based trust in prioritized default logic. In *Proceedings, The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence (AAAI'06, IAAI'06)*, pages 1345–1350, 2006. 153

D. E. Knuth. *Fundamental Algorithms.* Addison-Wesley Pub. Co, 3rd edition, 1997. 186

R. E. Korf. Generalized game trees. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence - Volume 1*, IJCAI'89, pages 328–333. Morgan Kaufmann Publishers Inc., 1989. 236

S. Kripke. A completeness proof in modal logic. *Journal of Symbolic Logic*, 24: 1–14, 1959. 59

Y. Lashkari, M. Metral, and P. Maes. Collaborative interface agents. In *National Conference on Artificial Intelligence*, pages 444–449. AAAI Press, 1994. 178

J. Leite and J. Martins. Social abstract argumentation. In *Proceedings of the Twenty-second International Joint Conference on Artificial Intelligence (IJCAI'11)*, pages 2287–2292, 2011. 286

H. Li, N. Oren, and T. J. Norman. Probabilistic argumentation frameworks. In *Proceedings of the First International Workshop on Theory and Applications of Formal Argumentation (TAFA'11)*, pages 1–16, 2011. 186

N. Litvak and R. van der Hofstad. Large scale-free networks are not disassortative. *CoRR*, abs/1204.0266, 2012. 199

J. D. Mackenzie. Question-begging in non-cumulative systems. *Journal of Philosophical Logic*, 8(1):117–133, 1979. 74

P. McBurney and S. Parsons. Games That Agents Play: A Formal Framework for Dialogues between Autonomous Agents. *Logic, Language and Information*, 11(3):315–334, 2002. 28, 57, 58, 73, 77, 238

P. McBurney and S. Parsons. Dialogue games for agent argumentation. *Argumentation in Artificial Intelligence*, pages 261–280, 2009. 55, 56, 57, 58, 59, 63

S. Modgil. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, 173(9-10):901–934, 2009. 140

S. Modgil and M. Caminada. Proof theories and algorithms for abstract argumentation frameworks. In *Argumentation in Artificial Intelligence*, pages 105–129. Springer, 2009. 11, 19, 40, 41, 42, 43, 44, 45, 48, 59, 62, 82, 85, 86, 90, 91, 92, 94, 288, 296

S. Modgil and H. Prakken. Revisiting preferences and argumentation. In *Proceedings of the Twenty-second International Joint Conference on Artificial Intelligence (IJCAI'11)*, pages 1021–1026, 2011. 31, 51, 54

S. Modgil, F. Toni, F. Bex, I. Bratko, C. Chesevar, W. Dvok, M. Falappa, X. Fan, S. Gaggl, A. Garca, M. Gonzlez, T. Gordon, J. Leite, M. Moina, C. Reed, G. Simari, S. Szeider, P. Torroni, and S. Woltran. The added value of argumentation. In S. Ossowski, editor, *Agreement Technologies*, volume 8 of *Law, Governance and Technology Series*, pages 357–403. Springer Netherlands, 2013. 32, 47

A. M. Mood. *Introduction to the theory of statistics*. McGraw-Hill, Inc., 1963. 195

M. Morge, J. McGinnis, S. Bromuri, P. Mancarella, K. Stathis, and F. Toni. Argumentative agents for service-oriented computing. In *Multiagent Systems and Applications - Volume 1: Practice and Experience*, pages 217–255. Springer, 2013. 32

N. Oren and T. Norman. Arguing Using Opponent Models. In *Argumentation in Multi-Agent Systems*, volume 6057 of *Lecture Notes in Computer Science*, pages 160–174. Springer Berlin Heidelberg, 2010. 21, 22, 25, 26, 28, 66, 146, 148, 236, 237, 238

M. J. Osborne and A. Rubinstein. *A course in game theory*. The MIT Press, 1994. 145

H. Prakken. Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese*, 127:187–219, 2001. 103

H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Logic and Computation*, 15(6):1009–1040, 2005. 21, 23, 75, 84, 98, 103, 104, 140

H. Prakken. Formal systems for persuasion dialogue. *Knowledge Engineering Review*, 21:163–188, 2006. 20, 21, 22, 24, 28, 35, 64, 72, 73, 74, 78, 91

H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124, 2010. 19, 22, 23, 24, 37, 46, 48, 49, 52, 53, 54, 62, 209, 258

H. Prakken and G. Sartor. Argument-based logic programming with defeasible priorities. *Journal of Applied Non-classical Logics*, 7(1):25–75, 1997. 94

H. Prakken and G. Vreeswijk. Logics for defeasible argumentation. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 4 of *Handbook of Philosophical Logic*, pages 219–318. Springer Netherlands, 2002. 18, 288

I. Rahwan and K. Larson. Mechanism design for abstract argumentation. In *Proceedings of the Seventh International Joint Conference on Autonomous Agents and Multi-Agent Systems - Volume 2*, pages 1031–1038, 2008. 21, 22, 26

R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1-2):81–132, 1980. 46

M. Richardson, R. Agrawal, and P. Domingos. Trust management for the semantic web. In *Proceedings of the Second International Semantic Web Conference (ISWC'03)*, pages 351–368, 2003. 153

T. Rienstra, M. Thimm, and N. Oren. Opponent Models with Uncertainty for Strategic Argumentation. In *Proceedings of the Twenty-third International Joint Conference on Artificial Intelligence*, IJCAI'13, pages 332–338. AAAI Press, 2013. 21, 22, 28, 237, 291

R. Riveret, A. Rotolo, G. Sartor, H. Prakken, and B. Roth. Success chances in argument games: a probabilistic approach to legal disputes. In *Proceedings of the Twentieth annual Conference on Legal Knowledge and Information Systems: JURIX*, pages 99–108, 2007. 21, 22, 23, 24, 25, 26, 27, 28, 64, 293

R. Riveret, H. Prakken, A. Rotolo, and G. Sartor. Heuristics in argumentation: A game-theoretical investigation. In *Proceedings of the Second International Conference on Computational Models of Argument (COMMA'08)*, pages 324–335, 2008. 21, 22, 24, 25, 26, 27, 28, 64

M. Rovatsos, I. Rahwan, F. A. Fischer, and G. Weiß. Practical strategic reasoning and adaptation in rational argument-based negotiation. In *Proceedings of the Second International Workshop on Argumentation in Multi-Agent Systems (ArgMAS'05)*, pages 122–137, 2005. 21, 22, 30

J. Silverman. Hidden agendas and how to uncover them. *Medicine*, 33(2):27 – 29, 2005. 57, 153

M. Snaith and C. Reed. TOAST: online ASPIC+ implementation. In *Proceeding of the Fourth International Conference on Computational Models of Argument (COMMA'12)*, pages 509–510, 2012. 260, 295

K. Stathis. A game-based architecture for developing interactive components in computational logic. *Journal of Functional and Logic Programming*, 2000(5), 2000. 58

M. Thimm. Tweety: A comprehensive collection of java libraries for logical aspects of artificial intelligence and knowledge representation. In *Proceedings of the Fourteenth International Conference on Principles of Knowledge Representation and Reasoning (KR'14)*, 2014. 259, 295

P. Tolchinsky, S. Modgil, K. Atkinson, P. McBurney, and U. Cortés. Deliberation dialogues for reasoning about safety critical actions. *Autonomous Agents and Multi-Agent Systems*, 25(2):209–259, 2012. 32

B. Verheij. Dialectical argumentation with argumentation schemes: An approach to legal logic. *Artificial Intelligence and Law*, 11(2-3):167–195, 2003. 49

J. von Neumann. Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100:295–320, 1928. 21, 235

G. Vreeswijk. *Studies in Defeasible Argumentation*. PhD thesis, Free University Amsterdam, 1993. 49

G. Vreeswijk and H. Prakken. Credulous and sceptical argument games for preferred semantics. In *Proceedings of the Seventh European Workshop on Logics in Artificial Intelligence*, JELIA'00, pages 239–253, 2000. 19, 39, 40, 59

W. E. Walsh, R. Das, G. Tesauro, and J. O. Kephart. Analyzing complex strategic interactions in multi-agent systems. In *AAAI-02 Workshop on Game-Theoretic and Decision-Theoretic Agents*, pages 109–118, 2002. 21, 22, 26, 29

D. Walton. Examination dialogue: An argumentation framework for critically questioning an expert opinion. *Journal of Pragmatics*, 38(5):745 – 777, 2006. 56

D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue*. SUNY Press, 1995. 55, 56, 69

Y. Wang and M. P. Singh. Trust representation and aggregation in a distributed agent system. In *Proceedings of the Twenty-first National Conference on Artificial Intelligence (AAAI'06)*, pages 1425–1430, 2006. 153

Y. Tang, K. Cai, E. Sklar, P. McBurney, and S. Parsons. A system of argumentation for reasoning about trust. *Proceedings of the Eighth European Workshop on Multi-Agent Systems*, 2010. 152, 153